

Deugdelijke peilingen

De weergegeven opvattingen zijn die van de auteur(s) en komen niet noodzakelijkerwijs overeen met die van het CBS

2013 | 10

Deugdelijke peilingen

Inhoud

1. Wat is een peiling?	5
2. Peilingen door de eeuwen heen	13
3. Wat gaat u eigenlijk onderzoeken?	23
4. De vragenlijst	31
5. Het trekken van een steekproef	47
6. Het verzamelen van gegevens	67
7. Controle en correctie	79
8. Schattingen maken	87
9. Het probleem van de non-respons	111
10. Analyse van de gegevens	121
11. Publicatie van de resultaten	139
12. Een checklist voor peilingen	145
Literatuur	159
Bronvermelding	161
Register	163

Wat is een peiling?

1.1 Survey-onderzoek, enquêtes, peilingen ...

Voelt u zich wel eens onveilig? Heeft u op dit moment een betaald baan? Als er vandaag verkiezingen zouden zijn voor de Tweede Kamer, op welke partij stemt u dan? Vindt u dat de bevoegdheden van het Europese Parlement moeten worden uitgebreid? Hoe is uw financiële situatie op dit moment? Hoe is over het algemeen uw gezondheid? Dit zijn allemaal vragen die u voorgelegd kunt krijgen als u meedoet aan een opiniepeiling.

Er zijn verschillende namen voor dit soort onderzoek in omloop. In wetenschappelijke kringen heeft men het vaak over survey-onderzoek. Overheidsinstanties als het Centraal Bureau voor de Statistiek (CBS) noemen het ook wel enquêtes. Een voorbeeld is de Enquête Beroepsbevolking (EBB). Marktonderzoekbureaus hebben het vaak over peilingen. Typische voorbeelden zijn de Politieke Barometer van Ipsos, de politieke peilingen van Maurice de Hond en het EénVandaag Opiniepanel van de TROS en de AVRO.

Peilingen zijn niet wezenlijk anders dan surveys of enquêtes. Toch zijn er in de praktijk wel wat verschillen. Peilingen zijn vaak klein en snel. Er doen hooguit een paar duizend mensen aan mee en het aantal vragen is beperkt. Het gaat vooral om een snelle indicatie van de publieke opinie over een actueel onderwerp. Peilingen volgen de actualiteit op de voet. Surveys en enquêtes zijn vaak groter en minder snel. Meer mensen krijgen een uitnodiging om aan het onderzoek deel te nemen. De vragenlijsten zijn groter en soms ook ingewikkelder. De onderzoekers zullen zich inspannen om toch vooral correcte uitspraken te kunnen doen. Dat kan meer tijd kosten. Daarom kan het enige tijd duren voordat de uitkomsten beschikbaar komen.

Wij zullen het vooral hebben over peilingen, maar daarbij dient u te bedenken dat alles wat we hier bespreken, ook van toepassing is op surveys en enquêtes.

We omschrijven een *peiling* als een vorm van onderzoek waarbij we gegevens verzamelen door het stellen van vragen aan mensen. Dat doen we omdat we meer te weten wilt komen over het gedrag of de mening van een grote groep mensen. Het kan bijvoorbeeld gaan om alle stemgerechtigden, alle leden van een vakbond, of alle leraren in het voortgezet onderwijs.

Het voorleggen van een reeks vragen aan elk lid van de groep kan een hele klus zijn, vooral als de groep groot is. Het zal dan veel tijd en geld kosten om alle informatie binnen te krijgen. U kunt de werkzaamheden en kosten binnen de perken houden door slechts een *steekproef* uit die groep te onderzoeken. De vraag is dan wat u wel en niet kunt doen met de gegevens die u in de steekproef hebben verzameld. Kunt u daarmee iets zinnigs zeggen over de groep als geheel? Ja, dan kan, maar er zitten wel een paar haken en ogen aan. De belangrijkste voorwaarde is dat die steekproef

op een correcte manier moet zijn getrokken. We zullen in deze leidraad uitleggen wat een goede steekproef is en hoe u daaruit juiste conclusies kunnen trekken.

Een goede peiling is niet alleen een kwestie van een goede steekproef trekken. Er komt nog wel wat meer voor kijken. Wat u met een peiling probeert te doen is het meten van allerlei zaken bij de deelnemers aan het onderzoek. Als u het gewicht van iemand willen weten, dan kunt u hem op een weegschaal zetten. En wilt u weten hoe hoog de bloeddruk van iemand is, dan kunt u die meten met een bloeddrukmeter. Voor het meten van meningen en gedragingen van personen zijn echter geen instrumenten beschikbaar. We moeten het daarom doen met het stellen van vragen. Daarbij kan veel misgaan waardoor het gegeven antwoord niet het juiste antwoord is. En als u verkeerde antwoorden krijgen op uw vragen, dan trekt u vervolgens ook verkeerde conclusies uit uw onderzoek. Daarom is het maken van een goede vragenlijst een zeer belangrijk onderdeel van het uitvoeren van een peiling. Ook dat komt in deze leidraad aan de orde.

Met een peiling kunt u op wetenschappelijk verantwoorde wijze betrouwbare informatie verzamelen over een grote groep, ook al hebben lang niet alle leden van die groep de vragenlijst ingevuld. Daar komt echter nog wel het nodige bij kijken. U moet zich aan allerlei principes en richtlijnen houden. En die kunnen het onderzoek best wel eens wat ingewikkelder maken. Maar dat is de prijs die u betaalt voor een deugdelijke peiling. Het is wel eens verleidelijk om je niet aan die richtlijnen te houden. Ook dan krijg je uitkomsten, maar het is de vraag wat die waard zijn. Er wordt tegenwoordig heel wat onderzocht. Er zijn goede peilingen en er zijn slechte peilingen. Het kaf is soms moeilijk van het koren te scheiden. We hopen dat deze leidraad u daarbij een handje kan helpen.

De titel van deze publicatie is 'Deugdelijke peilingen'. Met deugdelijk bedoelen we dat een peiling tegelijk betrouwbaar en valide moet zijn. *Betrouwbaar* betekent dat bij herhaald uitvoeren van dezelfde peiling (onder dezelfde omstandigheden) de uitkomsten hetzelfde moeten zijn. De peiling moet dus een *stabiel* meetinstrument zijn. *Valide* betekent dat de peiling werkelijk meet wat hij beoogt te meten. De uitkomsten moeten *geldig* zijn. Ze mogen je niet op het verkeerde been zetten.

Deze leidraad is bedoeld voor twee groepen mensen. De eerste groep bestaat uit zij die om wat voor reden dan ook met peilingen van anderen in aanraking komen. Misschien bent u journalist en krijgt u een persbericht onder ogen ('Uit onderzoek is gebleken dat ...') en vraagt u zich af of het beschreven onderzoek nu eigenlijk wel iets voorstelt. Of misschien bent u een bestuurder of beleidsmedewerker en wordt u geconfronteerd met de uitkomsten van een onderzoek waarvan u zich afvraagt hoe u die op waarde moet schatten.

De tweede groep bestaat uit zij die overwegen om zelf een peiling op te zetten en uit te voeren, maar die geen idee hebben hoe ze dat moeten doen. Misschien bent u van een buurtcomité en wilt u weten hoe buurtbewoners denken over bepaalde verkeersmaatregelen. Of misschien bent u een lokale journalist en wilt u peilen hoe de partijen het doen in de aanloop naar de gemeenteraadsverkiezingen

1.2 Een peiling in vogelvlucht

Deze leidraad neemt u stap voor stap mee door het hele proces van het opzetten en uitvoeren van een peiling. U kunt de leidraad echter ook goed gebruiken bij het beoordelen van de kwaliteit van onderzoek dat door anderen is uitgevoerd. Hoofdstuk 12 van deze handleiding bevat hiervoor een checklist.

In deze handleiding leggen we u uit hoe een deugdelijk peiling in elkaar zit. We zullen voorbeelden gebruiken om allerlei aspecten te illustreren. Die voorbeelden staan steeds in blauwe kaders als dit. Ze zijn niet essentieel voor de loop van het verhaal. U kunt ze dus eventueel overslaan zonder de draad van het verhaal kwijt te raken.

Voor het berekenen van de uitkomsten van een peiling hebt u soms enige wiskunde nodig. In de lopende tekst vermijden we wiskundige formules. Als u echter de precieze details wilt weten, dan kunt u die vinden in gele kaders als dit.

De principes voor het uitvoeren van betrouwbare peilingen zijn gebaseerd op de uitkomsten van wetenschappelijk onderzoek. We noemen dit onderzoeksterrein ook wel *survey-methodologie*. Er zijn hierover dikke boeken met wiskundige formules volgeschreven. In deze handleiding houden we het simpel en beperken we ons tot de basisprincipes. Als u dieper in de stof wilt duiken, dan kunnen we u verwijzen naar, bijvoorbeeld, Bethlehem (2009a).

Het opzetten en uitvoeren van een peiling is een proces dat u in een aantal stappen doorloopt. In deze paragraaf geven we een globaal overzicht hiervan. In de rest van de leidraad werken we dat verder uit.

Het begint allemaal met de opzet van het onderzoek. Daarin moet u precies vastleggen welke groep mensen u gaat onderzoeken, wat u bij die groep gaat onderzoeken en hoe u het gaat onderzoeken.

De doelpopulatie

Een van de eerste zaken die u bij de opzet van uw onderzoek moet vastleggen, is de *doelpopulatie*. Dat is de doelgroep, de gehele groep van objecten (mensen, huishoudens, bedrijven, ziekenhuizen, scholen) die u wilt onderzoeken en waarover u uitspraken wilt doen met de gegevens die u in het onderzoek verzamelt.

De peildatum

Een peiling is altijd een momentopname. U doet uitspraken over de toestand van de doelpopulatie op een bepaald tijdstip. Op een ander tijdstip kan de situatie immers weer anders zijn. Dat tijdstip (de *peildatum*) moet voor iedereen duidelijk zijn. Bij het uitvoeren van het onderzoek zal het meestal niet mogelijk zijn om alle waarnemingen precies op die peildatum te doen. U zult het werk daarom uitsmeren over een periode die daar omheen gespreid ligt. Daarbij moet u er goed voor waken

dat u de toestand wilt weten op de peildatum en niet op de dag van enquêtering. Dat moet in de vragen tot uiting komen.

Variabelen

Een volgend aspect van de opzet van het onderzoek is het vastleggen van de *variabelen*. Dat zijn de eigenschappen van de objecten in de doelpopulatie die u wilt meten. Zo zou u zich bij een peiling over internetgebruik kunnen vragen of iemand een internetaansluiting heeft en zo ja, hoeveel uur hij of zij per dag bezig is op het internet. De variabelen die u zo meet, leveren allemaal stukjes informatie op die u later in de analyse kunt combineren. Zo ontstaat een beeld van wat er in de populatie aan de hand is.

Populatiekenmerken

U gaat de verzamelde gegevens gebruiken om uitspraken te doen over de doelpopulatie. Het komt erop neer dat u het gedrag en de structuur van de populatie, probeert te beschrijven in een aantal kerncijfers. Dergelijke cijfers noemen we *populatiekenmerken*. Voorbeelden van zulke grootheden zijn: het *percentage* huishoudens dat thuis een internetaansluiting heeft, het *gemiddelde* aantal uren per week dat iemand surft op het internet, en het *totaal* aantal fietsen dat in het afgelopen jaar in huishoudens is gestolen.

De vragenlijst

Wat de objecten in de doelpopulatie ook zijn (mensen, huishoudens, bedrijven, enz.), u krijgt uw gegevens door het stellen van vragen aan mensen die de objecten vertegenwoordigen. We noemen dat de *respondenten*. De gegevens moeten consistent en vergelijkbaar zijn. Daarom moet u de gegevens zo uniform en objectief mogelijk verzamelen. U moet de vragen aan alle respondenten op precies dezelfde wijze stellen. Daarvoor is een *vragenlijst* het aangewezen instrument. Het maken van een vragenlijst vereist grote zorgvuldigheid. Er kan veel mis gaan bij het formuleren van een vraag. Daardoor kunt u de respondenten vrij eenvoudig op het verkeerde been zetten, waardoor u de verkeerde antwoorden krijgt, met alle gevolgen van dien.

De methode van gegevensverzameling

Voor elk object in de steekproef moet iemand de vragenlijst invullen. Dit kunt u op allerlei manieren laten doen, en elke manier heeft zijn voor- en nadelen in termen van kosten en kwaliteit. Bij *schriftelijke enquêteren* stuurt u de vragenlijsten per post naar de respondenten, met daarbij het verzoek ze ingevuld terug te sturen. Dit is een relatief goedkope methode, maar de kwaliteit van de gegevens kan te wensen over laten. *Mondeling enquêteren* is meestal beter. Hierbij bezoeken enquêteurs de geselecteerde personen, gezinnen of bedrijven en proberen in een persoonlijk gesprek de vragen beantwoord te krijgen. Dit is echter wel een kostbare methode. Een compromis wordt geboden door *telefonisch enquêteren*. Ook hierbij zet u

enquêteurs in, maar nu nemen ze de interviews vanaf een centrale plek (call center) telefonisch af. Voor telefonisch interviewen zijn minder enquêteurs nodig dan voor mondeling interviewen. De belangrijkste reden is dat ze niet hoeven te reizen van het ene adres naar het andere.

De computer speelt tegenwoordig een belangrijker rol bij het verzamelen van gegevens. Veel onderzoeksbureaus hebben het enquêteren met papieren vragenlijsten vervangen door *computergestuurd enquêteren*. De vragenlijst staat niet meer op papier, maar zit in een computer. De kwaliteit van de zo verzamelde gegevens is meestal beter, en u kunt de gegevens ook sneller verwerken. De mondelinge vorm van computergestuurd enquêteren noemen we meestal CAPI (*Computer-Assisted Personal Interviewing*) en de telefonisch vorm CATI (*Computer-Assisted Telephone Interviewing*).

Voor de laatste jaren is een nieuwe vorm van enquêteren erg populair geworden. Dat is enquêteren via het internet. Bij *online enquêteren* staat de vragenlijst in elektronische vorm op een website. Deze vorm van enquêteren heeft aantrekkelijke kanten, maar er zijn ook lastige problemen. We komen daar nog op terug.

De keuze van het steekproefkader

Als u weet hoe u de gegevens gaat verzamelen, dan moet u nadenken over de manier waarop u de steekproef gaat trekken. U hebt daarvoor een lijst nodig waarin de hele populatie staat opgesomd. We noemen dat een *steekproefkader*. Voorbeelden zijn een papieren kaartenbak (de ledenadministratie van de tennisclub), een elektronische databank (het bevolkingsregister van een gemeente), een lijst met telefoonnummers (het telefoonboek) en een lijst met e-mailadressen (van de studenten van een universiteit).

U trekt een steekproef door (via loting) een selectie van objecten te maken uit het steekproefkader. Om een goede steekproef te kunnen trekken, is het van belang dat het steekproefkader de doelpopulatie precies afdekt. Is dat niet het geval, dan kan dit tot problemen leiden die de validiteit van de uitkomsten aantasten.

De keuze van het steekproefontwerp

Het is een fundamenteel principe van de survey-methodologie dat u steekproeven moet loten. Alleen dan kunt u op basis van een steekproef deugdelijk uitspraken doen over de hele doelpopulatie. Er zijn allerlei manieren om een steekproef te loten uit een populatie. Zo zijn er de *aselecte steekproef* (met gelijke kansen), de *systematische steekproef* en de *tweetrapssteekproef*. Bij uw keuze zullen enerzijds praktische aspecten een rol spelen (hoe zit het steekproefkader in elkaar, en wat zijn de kosten?) en anderzijds de gewenste nauwkeurigheid van de uitkomsten.

Een belangrijke vraag die altijd weer opkomt bij het opzetten van een peiling, is hoe groot de steekproef moet zijn. Daarop kunnen we geen simpel antwoord geven. Als uw erg precieze uitkomsten wilt hebben, dan moet u een grote steekproef trekken. En als u bereid bent genoegen te nemen met een iets minder grote precisie, dan kunt

u volstaan met een kleinere steekproef. Later leggen we uit hoe u de precisie van de uitkomsten (de schattingen) kunt berekenen. Dan kunnen we ook uitleggen hoe u de steekproefomvang moet uitrekenen uitgaande van een gewenste precisie.

Het veldwerk

Hebt u de methode van gegevensverzameling gekozen en is de steekproef getrokken, dan moet u de vragenlijsten ingevuld zien te krijgen. Bij mondeling enquêteren (al of niet computergestuurd) gaan de enquêteurs op pad. Bij telefonisch enquêteren gaan de enquêteurs bellen. Bij schriftelijk enquêteren verstuurt u de vragenlijsten naar de respondenten. En bij een online enquête zet u de vragenlijst op het internet.

En dan maar hopen dat alle personen in de steekproef de vragenlijst keurig invullen. Helaas is dit lang niet altijd het geval. Een vervelend verschijnsel is het optreden van *non-respons*. Er kunnen allerlei redenen zijn waarom mensen de vragenlijsten niet invullen: ze zijn niet thuis, ze weigeren hun medewerking, of ze zijn niet in staat de vragen te beantwoorden (bijvoorbeeld wegens taalproblemen). Vooral als heel specifieke groepen niet meedoen, kan *non-respons* leiden tot een vertekend beeld, en dus tot verkeerde conclusies. We noemen dat een *selectieve non-respons*. Krijgt u te maken met *non-respons* in uw onderzoek, dan zult u moeten nagaan in hoeverre deze selectief is. En als dat het geval is, dan moet u daarvoor corrigeren.

Controle en correctie

Na verloop van tijd beëindigt u het veldwerk. Hopelijk hebt u dan een heleboel ingevulde vragenlijsten. Die gegevens moet u dan verder verwerken. Hebt u een of andere vorm van computergestuurd enquêteren gebruikt, dan zitten de gegevens al in de computer. Hebt u echter met papieren vragenlijsten gewerkt, dan moet u alle gegevens eerst invoeren in de computer voordat u ze verder kunt analyseren.

Tijdens het uitvoeren van een peiling kan er van alles misgaan. Dit kan leiden tot fouten in de verzamelde gegevens. Daarom moet u altijd een proces van *controle en correctie* uitvoeren. Daarbij controleert u de gegevens en probeert u zo goed mogelijk de fouten in de gegevens op te sporen en te verbeteren. Dat is lang niet altijd een eenvoudige zaak, zeker niet als dat achteraf moet gebeuren en de respondent niet meer beschikbaar is om u te helpen bij de oplossing van de problemen.

Correctie voor non-respons

Vrijwel altijd krijgt u als onderzoeker te maken met *non-respons*. Dit verschijnsel doet zich voor als u van sommige personen in de steekproef de gewenste informatie niet krijgt (om wat voor reden ook). *Non-respons* is vaak selectief. Het zijn vooral specifieke groepen in de doelpopulatie die niet of slecht meedoen aan het onderzoek. Om te voorkomen dat u onjuiste conclusies trekt uit uw onderzoek, moet u hiervoor corrigeren. Een veel gebruikte techniek hiervoor is *wegen*. Daarbij kent u gewichten toe aan de respondenten. Dat doet u zo dat ondervertegenwoordigde groepen zwaarder meetellen en oververtegenwoordigde groepen minder zwaar. Zo krijgt u

toch weer een evenwichtige (representatieve) steekproef. Daarbij moeten we helaas wel opmerken dat soms de consequenties van non-respons zo ernstig zijn dat we de problemen onvoldoende kunnen oplossen door weging.

Schattingen maken

Na het verzamelen en bewerken van de gegevens is het moment aangebroken om de cijfers echt te gaan analyseren. Dat zal in de eerste plaats neerkomen op het schatten van een aantal populatiekenmerken. Eventueel kunt u de schattingen nog uitsplitsen voor verschillende groepen. Daarmee kunt u bijvoorbeeld de vraag beantwoorden of jongeren er anders over denken dan ouderen, of mannen een ander gedrag vertonen dan vrouwen, en of kleine bedrijven minder investeren dan grote bedrijven.

U dient hierbij steeds te beseffen dat u die populatiekenmerken niet exact kunt uitrekenen. U heeft immers alleen maar gegevens uit een steekproef van personen tot uw beschikking. Dus zullen uw schattingen een zekere mate van onzekerheid hebben. Als u de steekproef netjes hebt geloot, dan kunt u de *onzekerheidsmarges* uitrekenen. Zo kunt u aangeven hoe precies uw uitspraken zijn. Dit soort informatie is van belang voor andere mensen die de uitkomsten van uw onderzoek op hun waarde willen schatten.

Publicatie van de resultaten

De laatste fase van het onderzoek is die van de *publicatie* van de resultaten. Dat betekent het schrijven van een rapport. Daarin dient u verantwoording af te leggen over de manier waarop u uw onderzoek hebt uitgevoerd. Iedereen moet kunnen vaststellen of u zich aan de regels voor wetenschappelijk verantwoord survey-onderzoek hebt gehouden.

Uiteraard moet u de uitkomsten van uw onderzoek publiceren. Zeker als het rapport bedoeld is voor een breed publiek, dient u ervoor te zorgen dat die uitkomsten niet op een verkeerde manier worden geïnterpreteerd. Het is daarom belangrijk dat u het hele onderzoeksproces en de kwaliteit van de uitkomsten beschrijft. U legt dus uit hoe de gegevens zijn verzameld, hoe hoog de respons was, hoe voor non-respons is gecorrigeerd en waar mogelijk vermeldt u onzekerheidsmarges bij de cijfers.

2. Peilingen door de eeuwen heen

De mensheid is al van oudsher bezig met het verzamelen van statistische gegevens, overigens zonder het zo te noemen. Eigenlijk gebeurt het al sinds het begin van de menselijke beschaving. Het waren vooral koningen, keizers en andere heersers die dit soort gegevens nodig hadden voor het besturen van hun landen of rijken. Door bij al hun onderdanen gegevens te verzamelen, kregen ze een redelijk beeld van de staat van hun gebieden. Ze konden zo bijvoorbeeld bepalen hoeveel belasting ze konden heffen en hoeveel soldaten ze op de been konden brengen.

Dit soort peilingen van de staat van het land waren *integrale onderzoeken*. Elke onderdaan moest gegevens leveren. Het idee van het gebruik van steekproeven kwam niet op. Tot het einde van de 19^{de} eeuw bleef men integraal onderzoek doen. Dat ging in de vorm van volkstellingen. Daarbij ging het meestal niet alleen om het tellen van de bevolking, maar ook om het verzamelen van andere informatie over de toestand in de samenleving.

De 19^{de} eeuw was een periode van industrialisatie. De bevolking nam in veel landen toe. Er ontstonden grote steden. Samenlevingen worden ingewikkelder. Centrale overheden hadden steeds meer informatie nodig om alles in goede banen te leiden. Door al die ontwikkelingen werd het uitvoeren van integraal onderzoek steeds lastiger, tijdrovender en kostbaarder. Omstreeks 1895 kwamen de eerste ideeën op over onderzoek op basis van steekproeven. De discussie over de zinvolheid en betrouwbaarheid van dit soort onderzoek duurde een jaar of 40. In 1934 was het theoretisch fundament voor de steekproeftheorie echter gelegd. Essentieel element daarin is dat we steekproeven moeten loten. Ook vandaag de dag is de kanssteekproef nog steeds een onontbeerlijk onderdeel van de survey-methodologie, en dus ook van elke peiling. Alleen als u de steekproef loot, kunt u deugdelijke conclusies trekken uit uw onderzoek, en alleen dan kunt u aangeven hoe precies uw schattingen zijn.

In dit hoofdstuk beschrijven we de, soms moeizame, ontwikkeling van de moderne survey-methodologie.

2.1 Het verzamelen van statistisch gegevens door de eeuwen heen

Koningen, keizers en andere heersers hebben altijd statistische overzichten gebruikt voor het besturen van hun landen of rijken. Al 1000 jaar voor Christus gaven de heersers van China en Egypte opdracht om statistieken te maken. Ze gebruikten deze gegevens voornamelijk voor het heffen van belastingen en militaire zaken. Ook de Romeinse keizers organiseerden regelmatige tellingen van mensen en hun bezittingen. Ze gebruikten de verzamelde gegevens om de politieke status van de inwoners te bepalen en om hun militaire en financiële verplichtingen vast te stellen. Bekend is het verhaal van keizer Augustus die omstreeks het jaar 0 een volkstelling

uitschreef waarvoor Jozef en Maria naar Bethlehem moest reizen. Figuur 2.1 bevat een deel van het schilderij dat Pieter Brueghel hierover maakte.

Figuur 2.1. Volkstelling te Bethlehem (Pieter Brueghel, 1605-1610)



Bij al dit soort onderzoek ging het om integraal onderzoek. Iedereen in de populatie moest meedoen. Er was geen sprake van steekproeven. Het idee was nog niet opgekomen dat dat je ook wel bruikbare statistieken zou kunnen maken op basis van minder gegevens.

Een mooi voorbeeld van een integrale peiling, maar wel van wat latere datum, is het *Domesday Book*. Willem de Veroveraar gaf opdracht tot het samenstellen van dit boek nadat hij in 1086 Engeland had veroverd vanuit het Franse Normandië. Het *Domesday Book* was het resultaat van een integraal onderzoek van de bezittingen van de koning en zijn vazallen in Engeland. Zo werden gegevens verzameld over meer dan 13.000 dorpen en riddersgoederen. Voor elk graafschap kwamen meer dan 10.000 cijfers beschikbaar.

Figuur 2.2. Het Domesday Book



Om al die cijfers te kunnen verzamelen, liet de koning het land opdelen in regio's. In elke regio benoemde hij een commissie. De leden daarvan waren belangrijke edelen. Die commissies organiseerden in elke stad bijeenkomsten. Iedereen die informatie kon verstrekken, moest voor de commissie verschijnen. Die werkte met een vaste vragenlijst. Daarin werd, bijvoorbeeld, gevraagd wie de eigenaar van een stuk land was, hoeveel vrije mensen en slaven er waren, wat er aan oppervlakte bos en

grasland was, hoeveel molens en visvijvers er waren, wat de totale waarde van het gebied was, en wat de winstvooruitzichten waren.

Zo kreeg Willem de Veroveraar een zeer waardevol statistisch overzicht, waarmee hij niet alleen zijn eigen rijkdom kon vaststellen, maar ook die van zijn vazallen. Veel informatie uit dit bijna 1.000 jaar oude statistische overzicht is bewaard gebleven, en vormt nu nog steeds een waardevolle bron van informatie over de geschiedenis van Engeland.

Een ander historisch voorbeeld van het gebruik van dit soort onderzoek vinden we terug in het rijk der Inca's, dat zo vanaf de 15^{de} eeuw in Zuid-Amerika zijn hoogtepunt bereikte. Elk Incadistrict had een *quipucamayoc*. Dit was een soort statisticus. Hij verzamelde allerlei gegevens over zaken als het aantal onderdanen, het aantal huizen dat zij bewoonden, hoeveel lama's er op de weiden graasden en het aantal jonge mannen dat geschikt was voor het leger. De quipucamayocs legden dat allemaal vast op quipu's. Een *quipu* was een systeem van geknoopte koorden van diverse kleuren. Elke kleur stond voor een bepaald onderwerp, en knopen in de koorden de aantallen (volgens het decimale systeem). We kunnen de quipu zien als een vroege voorganger van een enquêteformulier.

Figuur 2.3. De Quipucamayoc



De quipu's werden niet alleen gebruikt voor het bijhouden van economische gegevens, maar ook voor wetenschappelijke doeleinden. Zo waren er quipu's die de kringloop van de planeten registreerden. Het rijk der Inca's is ten onder gegaan, en daarmee ook de statistiek die daar werd bedreven.

De eerste moderne volkstelling in Noord-Amerika vond plaats in 1666 in Canada. Jean Talon was Intendant (gouverneur) van Nieuw-Frankrijk (Nouvelle-France). Hij wilde weten hoe het gebied zich had ontwikkeld sinds de stichting van Québec in 1608. Hij registreerde van alle inwoners geslacht, leeftijd, burgerlijke staat en beroep. Er bleken op dat moment 3.215 mensen te wonen in Nieuw-Frankrijk.

Figuur 2.4. De volksteller komt langs in (1870)



De Scandinavische landen liepen voorop bij de volkstellingen in Europa. De eerste volkstelling in Zweden vond plaats in 1748. Zowel de staat als de kerk hadden er belang bij. De staat wilde weten hoeveel mannen er konden worden opgeroepen voor militaire dienst en de kerk wilde in de gaten houden hoeveel mensen het protestantse geloof aanhingen. De eerste volkstelling in Denemarken werd georganiseerd in 1769.

Den Dulk & Van Maarsseveen (1990) beschrijven de ontwikkelingen in Nederland. De eerste volkstelling vond plaats in 1795. Nederland stond toen onder Franse invloed. Het was de periode van Bataafse Republiek. Er was een nieuw gecentraliseerd bestuur en dat wilde nieuwe kiesdistricten maken. Daarvoor moest men weten hoeveel mensen er overal woonden.

2.2 Het ontstaan van steekproefonderzoek

In de jaren voor 1895 waren het voornamelijk de nationale statistische bureaus die zich bezig hielden met het verzamelen van statistische gegevens. Die bureaus deden altijd integraal onderzoek. Het trekken van steekproeven was taboe. Waarom zou je een steekproef trekken als het ook met een integrale telling kon? En bovendien was het ongepast om echte waarnemingen te vervangen door rekenkundige manipulaties. Het was een vorm van discriminatie om een groot deel van de mensen uit te sluiten van deelname aan een onderzoek. Ook leek het de statistici in die tijd een onmogelijke zaak een uitspraak over een hele bevolking te doen als je maar gegevens over een klein deel daarvan had.

Toch doen we het zo vaak: het vellen van een oordeel over een groter geheel door slechts een klein deel ervan te onderzoeken. Voor de hand liggende voorbeelden zijn de kok die een lepel soep proeft, de bierbrouwer die een slokje bier neemt en de keurmeester die op de kaasmarkt een stukje uit het midden van de kaas steekt om het te proeven. Volgens sommigen is het woord “steekproef” zelfs afkomstig uit de wereld van het kaasmaken. Sinds de middeleeuwen zijn er al kaasmarkten in Nederland. Daar stak de keurmeester met een speciale kaasboor een stukje uit de kaas om de kwaliteit ervan vast te stellen.

Door de eeuwen heen is er dus altijd al gebruik gemaakt van steekproeven, misschien wel zonder het te beseffen, en ook zonder het zo te noemen. Het wetenschappelijk onderbouwde gebruik van steekproeven is echter van veel recentere datum.

Een belangrijke doorbraak vond plaats in het jaar 1895. In dat jaar kwam het ISI (International Statistical Institute) bijeen in het Zwitserse Bern. Het was Anders Kiaer (1895, 1997), de directeur van het Noorse Statistische Bureau, die het daar aandurfde om een pleidooi te houden voor het gebruik van steekproeven. Hij betoogde dat je met zijn *Representatieve Methode* goede resultaten kon behalen. Daarmee bedoelde hij een onderzoek waarbij hij waarnemingen deed bij een groot aantal (maar lang niet alle) objecten. Deze objecten koos hij verspreid over de hele doelpopulatie, en wel zo dat ze in hun totaliteit een kopie op kleine schaal van de populatie vormden. Een middel om tot een dergelijke representatieve selectie te komen was de *evenwichtige steekproef*. Op grond van kenmerken die voor elk object in de populatie bekend waren, selecteerde hij personen op zo'n manier dat de verdeling van de kenmerken in de steekproef overeenkwam met die in de populatie. Kiaer zorgde er bijvoorbeeld voor dat de verhouding man-vrouw in de steekproef overeen kwam met de verhouding man-vrouw in de populatie. En ook selecteerde hij de personen in de stad en op het platteland in de juiste verhouding.

Bij de selectie van de steekproef maakte Kaier geen gebruik van loting. Hij zocht via looproutes net zolang personen bij elkaar tot hij een steekproef had met de gewenste samenstelling. We zouden dat nu een *quota-steekproef* noemen. We zullen verderop in dit hoofdstuk laten zien dat het je met dit soort steekproeven een ernstig risico loopt verkeerde conclusies uit je peiling te trekken.

Een probleem van de Representatieve Methode van Kaier was dat hij geen idee had hoe goed of slecht zijn schattingen waren. Andere statistici vonden dit een ernstig nadeel. Daarom was er jarenlang veel discussie over de toepassing van de Representatieve Methode in de praktijk.

Het was Arthur Bowley (1906, 1926) die voor het eerst een theorie ontwikkelde waarmee we de onzekerheid in de uitkomsten van een peiling konden kwantificeren. Hij stelde voor om steekproeven te loten. De samenstelling van de steekproef wordt dan helemaal door het toeval bepaald. Er kunnen geen systematische effecten optreden. Voordeel van deze aanpak is ook dat je dan allerlei resultaten uit de theorie van de kansrekening kunt toepassen. Bowley toonde bijvoorbeeld aan dat schattingen bij benadering een zogenaamde *normale verdeling* hebben. Vervolgens kon hij uitrekeningen hoe ver schatting en werkelijke (te schatten) waarde van elkaar af kunnen liggen.

De Poolse wetenschapper Jerzy Neyman bewees in 1934 een aantal fundamentele resultaten die de verdere ontwikkeling van de steekproeftheorie hebben bepaald. Zo introduceerde hij het *betrouwbaarheidsinterval* als instrument om de precisie van een schatting aan te geven. Dat betrouwbaarheidsinterval gebruiken we nog steeds als onzekerheidsmarge van de uitkomsten.

In zijn artikel rekende Neyman ook af met steekproeven die niet door loting zijn verkregen. In een empirisch onderzoek van gegevens uit een Italiaanse volkstelling liet hij zien dat gebruik van zulke steekproeven tot wezenlijk verkeerde schattingen kan leiden, ook al is de steekproef representatief naar een aantal achtergrondkenmerken. Neyman toonde in feite aan dat de kanssteekproef de enige wetenschappelijk verantwoorde manier is om op basis van een steekproef een conclusie te trekken over een hele populatie.

2.3 De eerste opiniepeilingen

In de loop van de 30-er jaren van de vorige eeuw was duidelijk geworden dat peilingen op basis van kanssteekproeven een geschikt middel zijn om betrouwbare en valide uitspraken over populaties te doen. Toch duurde het nog heel wat jaren voordat dit soort peilingen in de dagelijkse praktijk werden toegepast. In 1948 voerde het CBS een vergelijkend onderzoek uit naar een aantal verschillende steekproeftechnieken. Zie voor meer informatie CBS (1948). De studie toonde aan dat je op basis van een steekproef van 30.000 personen uit de populatie van 1,75 miljoen belastingbetalers nauwkeurige schattingen kunnen maken. In dit onderzoek rekende het CBS voor het eerst betrouwbaarheidsintervallen uit.

Zelfs in 1952 moest de toenmalige Directeur-Generaal Idenburg van het CBS nog een pleidooi houden voor het gebruik van steekproeven. In een artikel voor de Centrale Commissie voor de Statistiek legt hij omstandig uit (zie Idenburg, 1952) hoe wetenschappelijk verantwoord gebruik van steekproeven nauwkeurige schattingen van populatiekarakteristieken kan opleveren. Hij ging ook in op de kritiek dat steekproeven niet iedereen gelijk behandelen. Sommige mensen zitten immers wel in de statistiek en anderen niet. Hij beantwoorde deze kritiek door te stellen dat een kanssteekproef iedereen precies dezelfde kans geeft om bij te dragen aan een statistiek.

Het zijn vooral de nationale statistische bureaus die veel statistische gegevens verzamelen. Maar ze zijn niet de enige organisaties die dat doen. In de loop der tijd zijn er ook steeds meer opiniepeilingen gekomen. Daarin gaat het niet meer om het verzamelen van harde feiten over mensen en hun omstandigheden, maar vooral om meningen over politieke, economische en sociale onderwerpen. De ontwikkelingen in Amerika laten zien dat het een proces van vallen en opstaan is geweest voordat men tot de conclusie kwam dat kanssteekproeven essentieel zijn voor deugdelijke uitkomsten.

De geschiedenis van de opiniepeilingen in de VS gaat terug naar 1824. In dat jaar probeerden twee Amerikaanse kranten, de *Pennsylvanian* in Harrisburg en de *Star* in Raleigh, de politieke voorkeur van de kiezers te meten in de periode voor de presidentsverkiezingen van dat jaar. De kranten gebruikten wel steekproeven maar ze hadden weinig aandacht voor de manier waarop die steekproef tot stand kwam. Daarom viel er weinig zinnigs te zeggen over de juistheid van de uitkomsten. Zulke peilingen werden *straw polls* genoemd. Die uitdrukking komt van het boerenland. Boeren gooiden een handvol strootjes in de lucht om te zien van welke kant de wind

kwam. De kranten deden straw polls in the straten van de stad om te zien hoe de politieke wind waaide.

De Amerikaanse presidentsverkiezingen van 1936 waren een beslissend moment in de ontwikkeling van opiniepeilingen. In die verkiezingen namen de Democraat Franklin Roosevelt en de Republikein Alf Landon het tegen elkaar op. De leidende politieke peiler was in die tijd het tijdschrift *Literary Digest*. Het tijdschrift schreef 10 miljoen Amerikanen aan. De adressen hiervoor waren afkomstig uit lijsten van auto-eigenaren en uit telefoonboeken. Uiteindelijk vulden 2,4 miljoen Amerikanen de vragenlijst in.

Sinds 1935 was er een nieuwe peiler in Amerika. Dat was *George Gallup*. Hij beseftte dat je alleen goede voorspellingen kunt doen als de steekproef goed in elkaar zit. De steekproef moet representatief zijn. Hij maakte hiervoor gebruik van quota-steekproeven. Hij gaf instructies aan de enquêteurs over de aantallen mensen die ze in de verschillende groepen moesten enquêteren: zoveel vrouwen uit de middenklasse in de stad, zoveel mannen uit de lagere klasse op het platteland, enz. De omvang van de steekproef bedroeg bij Gallup 50.000. Daarmee was de steekproef van Gallup dus aanzienlijk kleiner dan die van *Literary Digest*.

Tabel 2.1 bevat de voorspellingen van beide peilers en de echte uitslag van de verkiezingen. *Literary Digest* zat er helemaal naast. Die peiler voorspelde dat Landon de verkiezingen zou winnen met 57%. Maar het werd Roosevelt met 61%. Gallup voorspelde de winnaar wel goed, al zat deze peiler er toch ook nog 5% naast.

Tabel 2.1. De Amerikaanse presidentsverkiezingen van 1936

Kandidaat	Voorspelling <i>Literary Digest</i>	Voorspelling Gallup	Verkiezings- uitslag
Roosevelt (D)	43%	56%	61%
Landon (R)	57%	44%	37%

Waarom was de voorspelling van *Literary Digest* zo slecht? Dat kwam omdat de samenstelling van de steekproef niet goed was. De adressen waren die van auto-eigenaren en telefoonbezitters. Dat waren in die tijd mensen met wat hogere inkomens. Die mensen stemden vooral Republikeins. Dus de Republikeinen waren oververtegenwoordigd in de steekproef, met als gevolg dat er teveel Landonstemmers in de peiling zaten.

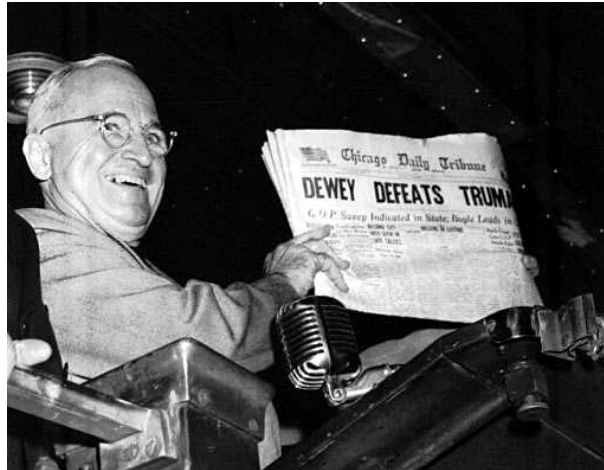
De quota-steekproeven van Gallup werkten in de praktijk ook niet altijd goed. Dat bleek bij de presidentsverkiezingen van 1948. Toen nam de Democraat Harry Truman het op tegen de Republikein Thomas Dewey. In tabel 2.2 staat de voorspelling van Gallup en de werkelijke verkiezingsuitslag.

Tabel 2.2. De Amerikaanse presidentsverkiezingen van 1948

Kandidaat	Voorspelling Gallup	Verkiezings- uitslag
Truman (D)	44%	50%
Dewey (R)	50%	45%

De steekproef van Gallup had dit keer een omvang van 3.250 personen. Op grond van de peiling voorspelde Gallup dat Dewey de verkiezingen zou winnen. Sommige kranten waren zo van overtuigd van de voorspelling van Gallup dat ze in hun vroege edities Dewey al tot winnaar verklaarden. Op de foto in figuur 2.4 houdt Truman triomfantelijk de krant omhoog met de foute voorspelling.

Figuur 2.5. De kranten verklaren Thomay Dewey tot winnaar



Gallup voorspelde dat Dewey 50% van de stemmen zou krijgen, en dat was 5% meer dan Dewey in werkelijkheid kreeg. Net als bij de verkiezingen van 1936, zaten er teveel Republikeinen in de steekproef van Gallup. Alleen leidde dat in 1936 niet tot een verkeerde voorspelling, omdat daarvoor het verschil tussen Roosevelt en Landon te groot was. In 1948 waren de verschillen tussen de kandidaten kleiner. De afwijking in de steekproef van Gallup zorgde er toen wel voor dat Gallup met de verkeerde voorspelling kwam.

De problemen met de voorspellingen van Gallup zijn voor een groot deel toe te schrijven aan het gebruik van quota-steekproeven. Dit soort steekproeven is niet gebaseerd op loting. Hij liet de enquêteurs porties mensen (quota) met bepaalde eigenschappen selecteren. Hij maakte zo zijn steekproeven representatief met betrekking tot variabelen als geslacht, leeftijd, opleidingsniveau en huidskleur. Maar dat betekent niet automatisch dat de steekproef ook representatief is met betrekking tot andere variabelen, zoals stemgedrag. Onderzoekers hebben inderdaad vastgesteld dat over een lange reeks van jaren de Republikeinen oververtegenwoordigd waren in dit soort quota-steekproeven.

Als gevolg van het fiasco van Gallup in 1948, besloot deze organisatie om af te stappen van het gebruik van quota-steekproeven. Vanaf dat moment werd alleen nog maar gebruik gemaakt van echte kanssteekproeven.

In Nederland zien we dat Unilever in 1934 het eerste marktonderzoekbureau opricht. Het heet Lintas (Lever's International Advertising Services). Een van de eerste projecten was een onderzoek naar het gebruik van het schuurmiddel VIM en leesgewoontes. Het project was zo succesvol dat Unilever in 1938 een nieuwe marktonderzoekbureau opzette: Instituut voor Huishoudelijk Onderzoek (IHO). In

1971 krijgt het een andere naam: Social and Marketing Research (SOCMAR), en in 1987 doopt Unilever het om in Research International Nederland. Na de Tweede Wereldoorlog onderzocht dit marktonderzoekbureau regelmatig het consumenten-gedrag met een panel van 600 huisvrouwen. Het is niet duidelijk hoe dit panel was opgezet en of het representatief was.

In 1940 ontstaat nog een ander onderzoeksbureau: de Nederlandse Stichting voor Statistiek (NSS). Het was een soort commerciële zuster van het CBS. De directeur van het CBS zat in de directie van het NSS. De belangrijkste activiteiten van het NSS waren marktonderzoek en opiniepeilingen.

In 1945 komt er nog een marktonderzoekbureau bij: het Nederlands Instituut voor de Publieke Opinie (NIPO). Het NIPO bracht in 1946 ook een tijdschrift uit: 'De Publieke Opinie'. Het eerste nummer legt uit dat je voor een goede peiling echt geen steekproef van 100.000 personen nodig hebt. Een omvang van 2.000 tot 10.000 is voldoende zolang je maar aan twee voorwaarden voldoet:

- Het moet een mondelinge enquête zijn en geen schriftelijke enquête;
- De steekproef moet representatief zijn met betrekking tot variabelen als inkomen, beroep, leeftijd en soms geloof.

Het bureau was tegen schriftelijke peilingen. Dan zouden er namelijk teveel mensen in de steekproef zitten met een hogere intelligentie en een hogere sociaal-economische positie. Zie ook NIPO (1946a).

Verkiezingspeilingen zijn altijd een mooie gelegenheid om te controleren of peilingen inderdaad doen wat ze moeten doen. De verkiezingen van 17 mei 1946 zijn een aardige illustratie daarvan. Tabel 2.3 vergelijkt de voorspelling van het NIPO met de werkelijke verkiezingsuitslag. De cijfers van het NIPO zijn gebaseerd op een peiling die twee weken voor de verkiezingen plaatsvond.

Tabel 2.3. De Nederlandse verkiezingen van 1946.

Partij	Verkiezings-uitslag	Voorspelling NIPO	Vershil
Partij van de Arbeid	28,3 %	33,9 %	5,6 %
Katholieke Volkspartij	30,8 %	29,5 %	1,3 %
Anti-Revolutionaire Partij	12,9 %	10,3 %	2,6 %
Christelijk Historische Unie	7,8 %	6,6 %	1,2 %
Partij van de Vrijheid	6,4 %	9,5 %	3,1 %
Communistische Partij Nederland	10,6 %	7,9 %	2,7 %
Protestantsche Unie	0,7 %	0,5 %	0,2 %
Staatkundig Gereformeerde Partij	2,1 %	0,9 %	1,2 %
Bellamy-Partij	0,2 %	0,8 %	0,6 %
Groep Lopes	0,1 %	0,1 %	0,0 %
Gemiddelde verschil			1,8 %

NIPO was niet echt blij met de uitkomsten. Een gemiddeld verschil van 1,8% was toch wel erg groot. Vooral de voorspelling voor de Partij van de Arbeid zat er met een verschil van 5,6% behoorlijk naast. Als verklaring verwees NIPO naar de gebeurtenissen op de avond voor de verkiezingen. Toen was er een toespraak van premier Schermerhorn op de radio waarin hij aankondigde het leger te gaan inzetten

om een staking te breken. Dat zou voor veel stemmers een reden zijn geweest om op het laatste moment van de Partij van de Arbeid over te stappen naar de Communistische Partij Nederland. Een meer gedetailleerde beschrijving is te vinden in NIPO (1946b).

Op 11 juni 1946 voerde NIPO de eerste telefonische peiling in Nederland uit. Aan een paar honderd huishoudens met telefoon in Amsterdam werden vragen gesteld over luisteren naar de radio. De mensen werden gebeld op een dinsdagavond tussen 20.00 uur en 21.30 uur. De resultaten van het veldwerk staan in tabel 2.4.

Tabel 2.4. De eerste telefonische enquête in Nederland

Luistert U op dit moment naar de radio?	Percentage
Ja	24 %
Nee	38 %
In gesprek	5 %
Geen antwoord	31 %
Heeft geen radio	2 %

Respondenten die naar de radio luisterden werd gevraagd naar welke programma ze luisterden. Het bleek dat 85% luisterde naar de “Bonte Dinsdagavondtrein”. Dat was in die tijd inderdaad een populair radioprogramma.

Over het algemeen gaven de marktonderzoekbureaus weinig informatie over de manier waarop ze hun steekproeven trokken. Een positieve uitzondering was misschien wel het NIPO dat in 1947 het belang van het loten van steekproef benadrukte, waarbij iedereen dezelfde kans moest hebben om in de steekproef te komen. Dat was ook de reden dat het NIPO de voorkeur gaf aan een mondeling enquêtes. Het idee was de arme mensen niet goed konden schrijven en bovendien ook niet waren geïnteresseerd in de onderwerpen van de peiling. Daarom zouden ze in een schriftelijke peiling een kleinere kans hebben om een ingevulde vragenlijst te produceren.

Het bovenstaande verhaal pretendeert niet een volledig beeld te geven van alle ontwikkelingen in Nederland op het gebied van peilingen. Het verhaal over het NIPO laat echter wel zien dat de opkomst van de peilingen een proces van vallen en opstaan was.

Het heeft na 1895 pakweg 50 jaar geduurd voordat de principes van de kanssteekproeven waren ingevoerd. Ondertussen is nu wel duidelijk geworden dat dit in de praktijk werkt en dat andere manieren voor het trekken van steekproeven tot grote problemen kunnen leiden. Meer over de opkomst van steekproefonderzoek is te vinden in Bethlehem (2009b).

3. Wat gaat u eigenlijk onderzoeken?

Voordat u begint met het uitvoeren van een peiling, moet u eerst een aantal zaken op een rij zetten. Daarvoor moet u de volgende vragen beantwoorden:

- Wie ga ik onderzoeken (de doelpopulatie)?
- Op welk moment wil ik de toestand in kaart brengen (peildatum)?
- Wat wil ik weten (populatiekenmerken)?
- Wat ga ik daarvoor meten (variabelen)?

In dit hoofdstuk leggen we uit hoe u deze vragen moet beantwoorden.

3.1 De doelpopulatie

De eerste stap in uw onderzoek is vaststellen welke groep objecten (mensen, huishoudens, bedrijven, scholen, enz.) u gaat onderzoeken. Die groep noemen we de *doelpopulatie*. Het is de hele groep van objecten waarover u in het onderzoek uitspraken wilt doen. Dit is dus ook de groep waaruit u de steekproef trekt.

Een goede omschrijving van de doelpopulatie is van groot belang en vereist daarom zorgvuldige overweging. Voor alle objecten die u 'in het veld' kunt tegenkomen, moet u zonder meer kunnen vaststellen of ze wel of niet tot de doelpopulatie behoren. Als u dit niet goed vastlegt, dan kan dit bij het veldwerk leiden tot vergissingen en problemen. Het kan gebeuren dat u ten onrechte objecten onderzoekt die niet in de doelpopulatie thuishoren, of u negeert objecten die er wel in thuishoren. Het is denkbaar dat u daardoor verkeerde conclusies trekt uit uw onderzoek.

Voorbeeld 3.1. ICT-onderzoek

Jaarlijks voert het Centraal Bureau voor de Statistiek (CBS) een onderzoek uit waarin informatie wordt verzameld over het gebruik van ICT-apparatuur en het internet door personen en huishoudens. Dit onderzoek vindt ook plaats in de andere lidstaten van de EU.

In dit onderzoek worden twee doelpopulaties onderscheiden. De eerste doelpopulatie bestaat uit alle personen van 12 t/m 74 jaar die in Nederland wonen, exclusief personen in inrichtingen en tehuizen.

De tweede doelpopulatie bestaat uit huishoudens. Het gaat om de huishoudens met ten minste één persoon in de leeftijd van 12 t/m 74 jaar. Die persoon moet in Nederland wonen en niet in een inrichting van een tehuis wonen.

Voorbeeld 3.2. Luisteronderzoek

Nederland kent bijna 300 lokale omroepen. Die omroepen voeren vrij regelmatig een luisteronderzoek uit. Zo kunnen ze een beeld krijgen van hoeveel mensen er naar de omroep luisteren en naar welke programma's ze luisteren.

Wat is de doelpopulatie bij een luisteronderzoek? De meeste lokale omroepen zenden uit voor één gemeente. Dan zal de doelpopulatie in principe bestaan uit de inwoners van die gemeente. Dat dient u echter nog wel nader te preciseren. Gaat het om alle mensen die in het bevolkingsregister van de gemeente staan ingeschreven? Of ook om mensen die er tijdelijk wonen? En wat te zeggen van mensen die niet in de gemeente wonen, maar er wel werken (en misschien op hun werk naar de radio luisteren)?

Verder zou u nog kunnen denken aan een minimum leeftijd. Immers, baby's zullen niet (bewust) naar de radio luisteren. Het is bovendien lastig om een vragenlijst voor te leggen aan zeer jonge kinderen. Bij veel luisteronderzoek gaat het daarom alleen om mensen vanaf een zekere leeftijd. Bedenk daarbij wel dat veel lokale omroepen ook programma's voor kinderen uitzenden. Als de omroep meer wil weten over het luisteren naar deze programma's, dan zult u die leeftijdsgroep toch ook in het onderzoek moeten betrekken.

Het is niet ongebruikelijk bij dit soort luisteronderzoek dat de doelpopulatie bestaat uit alle personen die in de gemeente wonen, en die minimaal 13 jaar zijn.

Doelpopulatie

In wiskundige termen is de doelpopulatie een eindige verzameling U die bestaat uit N objecten: $U = \{1, 2, \dots, N\}$. Hierin is N de omvang van de doelpopulatie. De nummers 1 t/m N duiden de volgnummers van de objecten in de doelpopulatie aan.

3.2 De peildatum

Veel onderzoek is erop gericht de toestand van de doelpopulatie op een specifiek moment in kaart te brengen. Dit moment noemen we de *peildatum*. In de ideale situatie trekt u de steekproef op de peildatum en ook op die peildatum laat u alle vragenlijsten invullen. Helaas is dit in de praktijk meestal niet mogelijk. Daarom zult u de peiling uitvoeren in periode om de peildatum heen.

U moet de steekproef trekken voordat het veldwerk van start gaat. Die steekproef zult u dus trekken op een eerder tijdstip dan de peildatum. U trekt de steekproef dan in feite uit een andere doelpopulatie. Het is de doelpopulatie zoals die eruit ziet op de *trekkingsdatum* en niet op de peildatum. Er kan van alles gebeuren tussen trekkingsdatum en peildatum. Zo kunnen nieuwe personen toetreden tot de doelpopulatie omdat ze ondertussen zijn gaan voldoen aan de definitie van de doelpopulatie. Die nieuwe personen zult u dus nooit in de steekproef trekken. Het

omgekeerde kan ook gebeuren: personen verdwijnen tussen de trekingsdatum en de peildatum uit de doelpopulatie, bijvoorbeeld omdat ze zijn overleden. Deze personen horen niet thuis in de doelpopulatie op de peildatum. U moet ze dus negeren en niet enquêteren.

De conclusie is dat u altijd zorgvuldig moet nagaan of de personen in de steekproef ook inderdaad op de peildatum tot de doelpopulatie behoren. Dit kan vooral lastig zijn als er non-respons optreedt. U moet dan proberen vast te stellen of de non-respondenten tot de doelpopulatie behoren (zodat u ze alsnog moet proberen te interviewen) of dat de non-respondenten toch niet tot de doelpopulatie behoort (zodat u ze verder kunt negeren).

Voorbeeld 3.3. Problemen met de peildatum

Stel, u wilt onderzoek doen onder de bevolking van een bepaalde gemeente en de toestand op 1 mei in kaart brengen. Omdat u niet alles op één dag kunt doen, trekt u de steekproef op een eerder tijdstip, zeg 1 april. Het veldwerk van het onderzoek laat u uitvoeren in een periode van twee weken, in de week voor 1 mei en in de week na 1 mei.

Stel, de enquêteur komt in de week voor 1 mei bij iemand aan de deur, en de desbetreffende persoon blijkt te zijn verhuisd naar een andere gemeente. Wat is er nu aan de hand? Deze persoon behoort niet tot de doelpopulatie. Immers, het gaat om de situatie op 1 mei, en op dat moment behoorde deze persoon niet meer tot de doelpopulatie. Dit is geen probleem. U kunt de persoon gewoon negeren.

Anders is de situatie als de enquêteur in de week na 1 mei aan de deur komt en de desbetreffende persoon blijkt na 1 mei verhuisd te zijn. Dan behoorde deze persoon dus wel tot de doelpopulatie en is er in feite sprake van non-respons. Om een mogelijke selectiviteit in het onderzoek tegen te gaan, moet u deze persoon opsporen en alsnog interviewen.

3.3 De variabelen

Aan de objecten in de steekproef gaat u metingen verrichten door het stellen van vragen. Zo krijgt u de gegevens die u nodig hebt voor uw analyse en voor het uiteindelijk trekken van conclusies over de toestand van de doelpopulatie. U kunt allerlei eigenschappen van objecten meten. Zo'n eigenschap noemen we een *variabele*. Die naam is zo gekozen omdat de eigenschap voor elk object anders kan zijn.

Voorbeelden van eigenschappen zijn de lengte van een persoon, het inkomen van een huishouden, het aantal werknemers van een bedrijf, de mestproductie van een boerderij en het aantal leerlingen op een school.

We maken onderscheid tussen kwalitatieve variabelen en kwantitatieve variabelen. Een *kwalitatieve variabele* wordt ook wel een *categorische variabele* genoemd. Hij verdeelt de doelpopulatie in groepen (categorieën). We kunnen niet rekenen met de

waarden van een dergelijke variabele. Het zijn slechts labels voor de verschillende groepen. We kunnen alleen vaststellen of objecten wel of niet tot dezelfde groep behoren. Voorbeelden van kwalitatieve variabelen zijn de geloofsovertuiging van een persoon (Rooms-katholiek, Protestant, Islam, Hindoe, enz.) en de bedrijfstak waarin een bedrijf actief is (landbouw, industrie, bouw, dienstverlening, enz.).

Een *kwantitatieve variabele* meet de omvang, waarde of duur van iets. Met die waarden kunnen we op zinvolle wijze rekenen. Er is ook altijd sprake van een meeteenheid. Voorbeelden van een kwantitatieve variabelen zijn het gewicht van een persoon (in kilogram), de ouderdom van een woning (in jaren) of de winst van een bedrijf (in 1000 euro's). We noemen een kwantitatieve variabele een *continue variabele* als hij binnen een bepaald interval elke mogelijke waarde kan aannemen. Een voorbeeld is de winst van een onderneming in een bepaald jaar. We noemen een kwantitatieve variabele een *discrete variabele* als hij aantallen meet. Het gaat om tellingen. De waarden kunnen dan alleen gehele getallen zijn. Voorbeeld: het aantal vrachtauto's dat een transportbedrijf heeft.

Met de waarden van kwantitatieve variabelen kunnen we zinvolle berekeningen uitvoeren zoals het bepalen van het gemiddelde waarde in de steekproef of de totale waarde in de steekproef.

Een speciaal type variabele is de variabele die de aan- of afwezigheid van een bepaalde eigenschap meet. Heeft een object die eigenschap, dan geven we dit aan met de waarde 1. Heeft het object die eigenschap niet, dan geven we dit aan met de waarde 0. Een dergelijk variabele noemen we een *indicatorvariabele*. Voorbeelden zijn het wel of niet hebben van internet en het wel of niet hebben van een betaalde baan.

We kunnen een indicator zien als een kwalitatieve variabele, want hij verdeelt de doelpopulatie in twee groepen: objecten die de desbetreffende eigenschap wel hebben en objecten die de eigenschap niet hebben. We kunnen de indicatorvariabele ook zien als een kwantitatieve variabele, want we kunnen rekenen met de waarden. Als we alle waarden in de steekproef optellen, krijgen we het aantal objecten in de steekproef met die eigenschap. Als we het gemiddelde nemen, dan krijgen we de fractie objecten in de steekproef met die eigenschap. En als we die uitkomst vermenigvuldigen met 100, dan krijgen we het percentage objecten met die eigenschap.

We onderscheiden doelvariabelen en hulpvariabelen. De *doelvariabelen* zijn de variabele die meten waar we op uit zijn in ons onderzoek. Ze meten allerlei aspecten van het verschijnsel dat we onderzoeken. In een vakantieonderzoek zouden de doelvariabelen bijvoorbeeld kunnen meten of men op vakantie gaat, waar naar toe, hoe lang, met welk vervoermiddel en hoeveel geld men uitgeeft.

Doelvariabele

We geven een willekeurige doelvariabele aan met de letter Y . De waarden van deze variabele voor de objecten in de doelpopulatie zijn dan Y_1, Y_2, \dots, Y_N . Is Y bijvoorbeeld de doelvariabele die meet hoeveel uur per week men op het internet zit, dan is Y_1 het aantal uur voor persoon 1, Y_2 het aantal uur voor persoon 2, enz.

Voorbeeld 3.4. Variabelen in een luisteronderzoek

Er zijn bijna 300 lokale omroepen in Nederland. Veel van die omroepen doen regelmatig een luisteronderzoek. Dat levert inzicht op in hoeveel mensen luisteren en waarnaar ze luisteren. In een poging al die luisteronderzoeken te vergelijkbaar te maken, heeft de brancheorganisatie OLON in het verleden voorstellen gedaan voor de doelvariabelen die in een luisteronderzoek kunnen worden gemeten. Een aantal daarvan staan hieronder genoemd.

Variabele	Type variabele
Bekendheid met de lokale omroep	Indicator
Heeft men ooit wel eens geluisterd	Indicator
Reden waarom men niet luistert	Kwalitatief
Weekbereik (heeft men in een willekeurige week geluisterd)	Indicator
Dagbereik (heeft men op een willekeurige dag geluisterd)	Indicator
Hoeveel uur men op een willekeurige dag heeft geluisterd	Kwantitatief
Soort programma waarnaar men luistert	Kwalitatief
Algemeen oordeel over omroep	Kwantitatief

De doelvariabelen meten dus allerlei aspecten van het verschijnsel dat u onderzoekt. Vaak meet u echter nog meer variabelen die zo op het oog weinig of niets met de doelstellingen van het onderzoek te maken hebben. Deze variabelen noemen we *hulpvariabelen*. Het gaat hier meestal om achtergrondkenmerken van de objecten. Bij personen zijn de hulpvariabelen vaak demografische variabelen zoals geslacht, leeftijd, burgerlijke staat en opleidingsniveau. Bij bedrijven kan het bijvoorbeeld gaan om de omvang van het personeel en het type economische activiteit.

Hulpvariabele

We geven een willekeurige hulpvariabele aan met de letter X . De waarden van deze variabele voor de objecten in de doelpopulatie noteren we met X_1, X_2, \dots, X_N . Stelt X bijvoorbeeld de hulpvariabele leeftijd voor, dan is X_1 de leeftijd van persoon 1, X_2 de leeftijd van persoon 2, enz.

Hulpvariabelen bieden u de mogelijkheid om in de analyse verschillende groepen met elkaar te vergelijken. Het zou kunnen zijn dat mannen zich anders gedragen dan vrouwen. Of dat ouderen anders scoren op een variabele dan jongeren. Of dat hoger

opgeleiden een ander gedrag vertonen dan lager opgeleiden. Kortom, de analyse van uw gegevens zal meer opleveren als u er hulpvariabelen bij betreft.

Hulpvariabelen zijn ook nog om een andere reden belangrijk. Ze zijn namelijk nodig om te corrigeren voor de akelige effecten van non-respons. Daarover kunt u meer lezen in hoofdstuk 9.

3.4 Populatiekenmerken

In uw peiling verzamelt u gegevens bij een steekproef van objecten. Met die gegevens wilt u uitspraken doen over de doelpopulatie. Het komt erop neer dat u de toestand in de populatie, probeert te beschrijven in een aantal kerncijfers. Dergelijke cijfers noemen we *populatiekenmerken*. Het zijn grootheden die u zou kunnen uitrekenen als u alle waarden van de bijbehorende doelvariabele in de doelpopulatie kende.

Voor een kwantitatieve doelvariabele kunt u twee populatiekenmerken definiëren: het *totaal* en het *gemiddelde* van de waarden van de doelvariabele. Stel dat de doelpopulatie uit huishoudens bestaat en de doelvariabele is het aantal fietsen dat in het huishouden aanwezig is. Dan is het populatietotaal gelijk aan alle fietsen in de doelpopulatie. En als de doelvariabele het netto maandinkomen van een persoon is, dan zou het gemiddeld maandinkomen in de doelpopulatie een interessant populatiekenmerk kunnen zijn. Als u dan ook de hulpvariabele geslacht meet in het onderzoek, dan kunt u bijvoorbeeld nagaan of het gemiddeld netto maandinkomen van mannen en vrouwen verschilt.

Voor het bepalen van de nauwkeurigheid van uw schattingen is nog een andere populatiekenmerk van belang. Dat is de populatievariantie, of iets specifieker, de *aangepaste populatievariantie*. Dit kenmerk geeft aan hoeveel variatie er zit in de waarden van de doelvariabele. Heeft elk object dezelfde waarde, dan is de aangepaste populatievariantie gelijk aan 0. Naarmate er meer variatie zit in de waarden, is de waarde van de aangepaste populatievariantie groter.

Voor een indicatorvariabele kunt u drie populatiekenmerken berekenen: het totaal aantal objecten in de doelpopulatie met een specifieke eigenschap, de fractie objecten met die eigenschap en het percentage objecten met die eigenschap.

Voorbeeld 3.5. Populatiekenmerken voor een luisteronderzoek

Als u in een luisteronderzoek van een lokale omroep vraagt of men gisteren naar de omroep heeft geluisterd, dan meet u een indicatorvariabele met waarden 1 (wel geluisterd) en 0 (niet geluisterd). Het totaal van deze variabele is het aantal mensen dat gisteren heeft geluisterd (in de steekproef) en het gemiddelde is de fractie mensen dat heeft geluisterd. Als u die fractie vermenigvuldigt met 100, krijgt u het percentage mensen dat heeft geluisterd.

Populatiekenmerken voor een kwantitatieve variabele

Het *populatietotaal* van de kwantitatieve doelvariabele Y is gelijk aan

$$Y_T = Y_1 + Y_2 + \dots + Y_N = \sum_{k=1}^N Y_k .$$

Het *populatiegemiddelde* van de kwantitatieve doelvariabele Y is gelijk aan

$$\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{Y_T}{N} = \frac{1}{N} \sum_{k=1}^N Y_k$$

De *aangepaste populatievariantie* van de kwantitatieve doelvariabele Y is gelijk aan

$$S^2 = \frac{(Y_1 - \bar{Y})^2 + (Y_2 - \bar{Y})^2 + \dots + (Y_N - \bar{Y})^2}{N - 1} = \frac{1}{N - 1} \sum_{k=1}^N (Y_k - \bar{Y})^2 .$$

De aangepast populatievariantie is een soort gemiddelde van de kwadraten van de afstanden van de waarden tot het gemiddelde.

We hebben al eerder aangegeven dat we met kwalitatieve variabelen maar weinig kunnen rekenen. Er zijn daarom geen specifieke populatiekenmerken voor kwalitatieve variabelen. Wat we wel kunnen doen, is het tellen van aantallen objecten in de verschillende categorieën. En die aantallen kunnen we eventueel omwerken naar percentages. In feite hebben we dan de kwalitatieve variabele omgezet in een reeks indicatorvariabelen, waarbij er een indicatorvariabele is voor elke categorie.

Populatiekenmerken voor een indicatorvariabele

Als Y een indicatorvariabele is (met waarden 0 en 1), dan is het *populatietotaal*

$$Y_T = Y_1 + Y_2 + \dots + Y_N = \sum_{k=1}^N Y_k$$

gelijk aan het aantal objecten met de desbetreffende eigenschap.

Het *populatiegemiddelde*

$$\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{Y_T}{N} = \frac{1}{N} \sum_{k=1}^N Y_k$$

van de indicatorvariabele Y is gelijk aan de fractie objecten met die eigenschap.

Als we het percentage objecten in de populatie met een bepaalde eigenschap aanduiden met P , dan geldt voor dit *populatiepercentage*:

$$P = 100 \times \bar{Y} = 100 \times \frac{Y_1 + Y_2 + \dots + Y_N}{N} = 100 \times \frac{Y_T}{N}$$

De *aangepaste populatievariantie* is voor een indicatorvariabele Y gelijk aan

$$S^2 = \frac{N}{N-1} \times \frac{P}{100} \times \frac{100-P}{100}.$$

4. De vragenlijst

In uw peiling verzamelt u gegevens door het stellen van vragen aan respondenten. Die gegevens moeten een juist beeld geven van de respondenten, en bovendien moeten ze vergelijkbaar zijn voor de verschillende respondenten. Daarom is er behoefte aan een objectief meetinstrument. Dat meetinstrument is de vragenlijst.

Bij het ontwikkelen van de vragenlijst moet u grote zorg besteden aan de formulering van elke vraag en de beschrijving van het soort antwoord dat u verwacht. Immers, de vragenlijst is het meetinstrument waarmee u de gegevens verzamelt. Als dat meetinstrument niet goed werkt, dan zal dat doorwerken in het gehele onderzoek. U loopt dan het risico verkeerde conclusies te trekken.

Er wordt wel eens gezegd dat het ontwerpen van vragenlijsten meer een kunst dan een kunde is. Het is inderdaad lastig om harde regels te geven voor het maken van goede vragenlijsten. Niettemin heeft wereldwijde en jarenlange ervaring toch wel wat vuistregels opgeleverd. We zullen een aantal van die vuistregels in dit hoofdstuk beschrijven. In het eerste deel van dit hoofdstuk gaat het vooral om de formulering van de vragen. Het tweede deel gaat over de structuur van de vragenlijst, zoals de volgorde van de vragen, en de voorwaarden waaronder u ze moet stellen. In het derde deel gaan we in op het testen van vragenlijsten.

4.1 De formulering van de vraagtekst

Een van de belangrijkste aspecten van de vragenlijst is de formulering van de tekst van de vraag. Deze moet begrijpelijk zijn voor iedereen die hem in het onderzoek voorgelegd krijgt. Het zijn vaak hoger opgeleide personen die de vragenlijsten opstellen. Ze weten veel van het onderwerp van de peiling. Het gevaar bestaat dan dat ze teveel moeilijke woorden en teveel vakjargon gebruiken.

Stel begrijpelijke vragen

Een voorbeeld van een begrip dat voor wetenschappers relatief eenvoudig is te begrijpen, maar toch geen alledaagse kost is voor de gemiddelde Nederlander, is de snelheid waarmee veranderingen plaatsvinden. De volgende vraag illustreert dit:

Vindt u dat de prijzen van levensmiddelen op het ogenblik sneller stijgen, even snel stijgen of langzamer stijgen dan een jaar geleden?

Lang niet elke respondent heeft door wat u hier vraagt. Het gaat niet om het stijgen of dalen van de prijzen, maar om de snelheid waarmee de prijzen stijgen. Het is een lastige vraag en daarom ligt het gevaar op de loer dat hij verkeerd wordt begrepen.

Om de tekst van een vraag voor iedereen begrijpelijk te houden, moet u de tekst van de vraag niet te lang maken en simpele woorden gebruiken. Probeer de tekst op te schrijven in ‘gesproken taal’ en niet in ‘geschreven taal’. Stel u legt een respondent de volgende vraag voor:

Bent u tevreden over de recreatieve voorzieningen in uw woonplaats?

De kans is dan heel groot dat de respondent niet precies weet wat u bedoeld met 'recreatieve voorzieningen'. Als het mee zit, dan zal hij denken aan een zwembad, maar of hij ook zal denken aan zaken als de bibliotheek en het stadspark, valt te betwijfelen. In deze situatie is het beter om concreet te omschrijven wat u precies bedoelt.

Aardige indicatoren voor de moeilijkheidsgraad van een zin zijn het aantal lettergrepen en het gemiddeld aantal lettergrepen per woord. Bevat de tekst veel moeilijke woorden, dan zal dit gemiddelde relatief hoog zijn.

Het algemene advies is om de tekst van de vraag kort te houden. Er zijn echter situaties denkbaar waarin de tekst wat langer zou kunnen zijn. Het idee daarbij is om bij een mondeling of telefonisch interview de respondent wat meer tijd te geven om na te denken over het antwoord op de vraag.

Stel ondubbelzinnige vragen

Zelfs als een vraag in simpele taal is gesteld, dan betekent dat nog niet dat daarmee alle problemen zijn opgelost. De vraag moet ook ondubbelzinnig, maar voor één uitleg vatbaar, zijn. Een vraag als

Wanneer bent u van school afgegaan?

is zonder nadere toelichting voor allerlei uitleg vatbaar. Antwoorden als 'na mijn eindexamen', 'in 1974', 'toen ik ging trouwen', 'toen ik 18 was' en 'toen ik zwanger was' zouden allemaal kunnen. Als u, bijvoorbeeld, een datum wilt weten, vraag dan ook naar die datum.

Over het algemeen is het voor respondenten makkelijker om op feitelijke vragen antwoord te geven dan op opinievragen. Maar zelfs feitelijke vragen kunnen onduidelijk en verwarrend zijn. Als u aan een werkloze vraagt of hij op zoek is naar werk, dan kan dat van alles betekenen. Op zoek zijn naar werk kan neerkomen op zelf initiatief nemen en allerlei bedrijven aanschrijven, maar het kan ook betekenen dat hij de hele dag alleen maar uit het raam zit te staren. Het is zelfs denkbaar dat de vraag suggestief is. De werkloze zoekt eigenlijk helemaal niet naar werk, maar hij antwoordt dat hij dat wel doet omdat dit sociaalwenselijk gedrag is.

Bij het opstellen van een vraag dient u te beseffen dat de interpretatie die een respondent geeft aan een bepaalde terminologie, sterk afhangt van zijn referentiekader. Dat hoeft niet voor geïnterviewde en interviewer hetzelfde te zijn. Zo kan een simpele term als 'gezin' door iemand in enge zin worden gedefinieerd als de partners met hun kinderen, maar het kan ook heel breed worden gedefinieerd als alle personen in het huishouden, dus inclusief inwonende opa's en oma's. En als u aan de respondent vraagt hoeveel kamers er in de woning zijn, wat telt er dan allemaal mee? De bijkeuken? De badkamer? De hal?

Stel geen suggestieve vragen

De formulering van de vraag moet neutraal zijn. De respondent moet een kans krijgen zijn eigen mening te geven, zonder dat u hem daarbij probeert te beïnvloeden. U moet dus suggestieve vragen vermijden. Met een vraag als

U hebt toch zeker ook wel een iPad?

Wekt u de indruk dat het eigenlijk abnormaal is om geen iPad te hebben, en daarom zal een respondent aarzelen dat toe te geven. Woorden als ‘ook’, ‘toch’ en ‘wel’ kunt u beter vermijden. Ook is het onjuist om deskundigen aan te halen, omdat niet iedereen altijd even makkelijk een deskundigen zal durven tegen te spreken.

Stel geen dubbele vragen

Een veel gemaakte fout is in een vraag meer dan één onderwerp aan de orde te stellen, terwijl er maar één antwoord kan worden gegeven. De vraag

Is het u bekend dat de lokale omroep een website heeft, en dat daarop een samenvatting is te vinden van al het lokale nieuws?

bevat in feite twee vragen. Als iemand dan antwoord met ‘ja’, waarop geeft hij dan antwoord? Op de eerste vraag (*Is het u bekend dat de lokale omroep een website heeft?*) of op de tweede vraag (*Is het u bekend dat de website van de lokale omroep een samenvatting is te vinden van al het lokale nieuws?*) of op beide vragen? Om dit soort problemen te voorkomen, moet u een vraag als deze splitsen in twee aparte vragen.

Vermijd (dubbele) ontkenningen

Vragen zijn ook moeilijk te begrijpen als ze een ontkenning bevatten, of nog erger, een dubbele ontkenning. Bij de vraag

Gaat u liever niet naar een café waar niet gerookt mag worden?

zult u even heel diep moeten nadenken wat het antwoord ‘nee’ eigenlijk betekent.

Pas op met vragen over gevoelige onderwerpen

Ook al hebt u de vraag duidelijk en begrijpelijk geformuleerd, dan nog kan het zo zijn dat u het juiste antwoord niet krijgt. Een reden kan zijn dat een respondent een *sociaalwenselijke antwoord* geeft. Als een onderwerp nogal gevoelig ligt (abortus, crimineel verleden, gebruik van drugs, seksueel gedrag), dan zal de respondent eerder een algemeen geaccepteerd antwoord geven dan zijn ‘eigen’ antwoord. Daarom is het twijfelachtig of de antwoorden op een vraag als

Bezoekt u wel eens porno-websites op het internet?

een correct beeld zullen geven van het surf-gedrag van de respondenten.

Pas op met vragen die een beroep doen op de herinnering

De mens onthoudt sommige gebeurtenissen beter dan andere gebeurtenissen. Daarmee moet u rekening houden als u een vraag stelt over gebeurtenissen die in het verleden hebben plaatsgevonden. Bij een vraag als

Hoe vaak hebt u de afgelopen drie maanden een bezoek gebracht aan uw huisarts?

zal een aantal bezoeken worden vergeten. Dit kan in sommige situaties oplopen tot 30%. Door de periode waarop de vraag betrekking heeft, korter te maken (bijvoorbeeld een week in plaats van drie maanden), verhoogt u wel de betrouwbaarheid van de antwoorden, maar daar staat tegenover dat u minder informatie krijgt.

Speciaal bij vragen over belangrijke gebeurtenissen in het verleden, kan ook nog het verschijnsel *telescoping* optreden. Daarbij plaatsen de respondenten gebeurtenissen op een verkeerd moment in de tijd. Vaak lijken ernstige gebeurtenissen veel korter geleden gebeurd te zijn dan in werkelijkheid het geval was. Bij vragen over hoeveel maal een bepaalde gebeurtenis in het een bepaalde periode heeft plaatsgevonden kan telescoping dus tot problemen leiden. Belangrijke gebeurtenissen worden te vaak gerapporteerd en onbelangrijke gebeurtenissen te weinig.

Wat soms kan helpen bij vragen die een beroep doen op de herinnering, is het verwijzen naar markante gebeurtenissen. In plaats van te vragen naar de datum van een gebeurtenis, zou u ook kunnen vragen of de gebeurtenis plaats vond voor of na de kerstdagen, of voor of na de Elfstedentocht.

Vermijd hypothetische vragen

Nog moeilijker dan vragen over het verleden zijn *hypothetische vragen*. Als u een hypothetische vraag stelt, dan krijgt u ook een hypothetisch antwoord. Als u mensen vraagt wat ze zouden hebben gedaan als zich in hun leven bepaalde gebeurtenissen wel of niet hadden voorgedaan, dan geeft u hen een heel moeilijke opdracht. U vraagt veel verbeeldingskracht van een respondent die op een dergelijke vraag een serieus antwoord wil geven. Die respondent zal flink wat tijd nodig hebben om alle aspecten van zo'n hypothetische situatie te overdenken. Hypothetische vragen worden vaak gesteld in een poging iets meer te weten te komen over de houding en mening van de respondent over allerlei zaken. Het is echter zeer weinig bekend over wat zich afspeelt in de geest bij de beantwoording van zulke vragen. Het is dus maar zeer de vraag of de vraag meet wat u wilt meten.

Bij de formulering van vragen gaan we er vaak vanuit dat een bepaalde hoeveelheid algemene informatie bij de respondenten bekend zal zijn. In de praktijk kan dit erg tegenvallen. Bekend is een onderzoek van Gallup uit de veertiger jaren van de vorige eeuw waarin aan een steekproef van Amerikanen werd gevraagd de 10 staten die ze het best kenden op de kaart aan te wijzen. Van de Amerikanen met een lagere

school opleiding bleek slechts 4% hiertoe in staat. En Amerikanen met een collegeopleiding waren al niet veel beter. Van hen slaagde slechts 8% voor de opgedragen taak. De boodschap is duidelijk. Ga er niet zonder meer vanuit dat allerlei kennis wel aanwezig zal zijn, en leg zo nodig zaken uit. Nalaten hiervan kan tot problemen leiden bij vervolgvragen die met dit onderwerp te maken hebben.

4.2 Soorten vragen

Tot nog toe is alleen de tekst van de vraag aan de orde geweest. Het is echter ook belangrijk om zorg te besteden aan het soort antwoord dat men moet geven op de vraag. We kunnen verschillende soorten vragen onderscheiden. Die verschillende soorten vragen leiden tot verschillende soorten antwoorden. Hier beperken we ons tot open vragen, gesloten vragen en numerieke vragen

Open vraag

Het meest voor de hand liggende soort vraag is de *open vraag*. Daarbij accepteert u elke willekeurige tekst als antwoord. Open vragen hebben het voordeel dat de respondent spontaan en in zijn eigen woorden kan reageren. Er zijn echter ook nadelen. Zo is het niet uitgesloten dat de respondent bepaalde antwoordmogelijkheden over het hoofd ziet. Het volgende voorbeeld illustreert dit probleem. In een survey werd de volgende open vraag gesteld:

Welke weekbladen leest u?

Als u de vraag zo stelt, dan blijken veel mensen allerlei bladen over het hoofd te zien. Dat bleek vooral het geval te zijn voor omroepbladen. Dit zijn toch echt ook weekbladen. Als u er een lijst van bladen bij vermeldt, en daar zitten ook de omroepbladen bij, dan blijken ineens veel meer mensen omroepbladen te lezen.

Het stellen van een open vraag kan ook leiden tot onduidelijke antwoorden. In een luisteronderzoek van een lokale omroep kwam de volgende vraag voor:

Wat vindt u het belangrijkste aspect van de lokale omroep?
.....

Veel respondenten gaven als antwoord 'het lokale karakter'. Daarbij werd het niet duidelijk wat ze daar precies mee bedoelden. Bij verder vragen bleek dat het bij sommigen ging om de lokale nieuwsvoorziening, terwijl anderen vooral dachten aan de omroep als lokale vrijwilligersorganisatie.

Wat doet u met al die antwoorden op een open vraag? Statistische analyse van dit soort tekstuele informatie is erg lastig. Daarom is het beter om open vragen zoveel mogelijk te vermijden. U moet alleen open vragen gebruiken als het echt niet anders kan. Een voorbeeld is de vraag naar het dorp of de stad waar de respondent zijn laatste zomervakantie heeft doorgebracht:

In welk dorp of welke stad heeft u uw laatste zomervakantie doorgebracht?

.....

U zou natuurlijk de respondenten kunnen laten kiezen uit een lijst van alle mogelijke vakantiebestemmingen. Die lijst is wel erg lang. Het is vrijwel ondoenlijk om daarin de juiste locatie te vinden. Er zijn misschien ook wel mensen die niet meer weten in welke plaats ze geweest zijn. Dan kunnen ze bij een open vraag in ieder geval de streek of het land invullen. Lastig is ook dat de er schrijffouten in de antwoorden kunnen voorkomen. Hopelijk levert dat geen verwarring op.

Gesloten vraag, 1 antwoord

Bij een *gesloten vraag* legt u aan de respondent een lijst met mogelijke antwoorden voor. Deze moet dan één van de mogelijkheden in de lijst uitkiezen en aankruisen. Dit vereist natuurlijk wel dat alle mogelijke antwoorden in de lijst staan, zodat elke respondent een passend antwoord kan vinden. Toch moet u vermijden dat de respondent een erg lange lijst met mogelijkheden krijgt voorgelegd. De keuze is moeilijker voor de respondent naarmate de lijst langer is:

Welk vervoermiddel gebruikt u het meest bij verplaatsingen binnen de grenzen van uw woongemeente?

- Lopend
- Fiets
- Elektrische fiets
- Bromfiets
- Bromscooter
- Motor
- Auto
- Bus
- Tram
- Metro / lightrail
- Ander vervoermiddel

Bij mondelinge en telefonische interviews leest de enquêteur alle mogelijke antwoorden op. De respondent kan die nooit allemaal onthouden. De eerste antwoorden is hij al weer snel vergeten en daarom kiest hij een antwoorden ergens achteraan in de lijst. Deze voorkeur van een antwoord achteraan in de lijst noemen we het *recency effect*. Bij schriftelijke enquêtes en internet-enquêtes moet de respondent zelf de lijst met mogelijke antwoorden lezen. Dat gebeurt niet altijd even zorgvuldig. De respondent verliest al gauw de aandacht. Dat leidt daarom juist tot een voorkeur voor een antwoord vooraan in de lijst. Dit noemen we het *primacy effect*. Om dit soort problemen te voorkomen, is het verstandig om een zo kort mogelijke lijst van mogelijke antwoorden aan te bieden. Bij computergestuurd interviewen zou u ook nog kunnen overwegen om de volgorde van de mogelijke antwoorden door het toeval te laten bepalen.

U kunt gesloten vragen goed gebruiken voor het peilen van de mening van personen. De respondenten kunnen dan kiezen uit een beperkt aantal antwoorden. Het is niet ongebruikelijk hiervoor een zogenaamde 5-puntsschaal te hanteren:

Hoe tevreden of ontevreden bent u over het huidige kabinet?
<input type="checkbox"/> Zeer tevreden
<input type="checkbox"/> Tevreden
<input type="checkbox"/> Niet tevreden en niet ontevreden
<input type="checkbox"/> Ontevreden
<input type="checkbox"/> Zeer ontevreden
<input type="checkbox"/> Weet niet

Merk op dat bij deze vraag nog een zesde optie “Weet niet” is opgenomen. Die is uiteraard bedoeld voor mensen die echt helemaal geen mening hebben. Helaas wordt dit antwoord ook vaak gekozen door mensen die wel een mening hebben, maar die niet willen geven. Het is daarmee een soort ‘nooduitgang’.

Het blijft een dilemma of u “Weet niet” wel of niet als mogelijk antwoord moet opnemen in de lijst. Is “Weet niet” aanwezig als een van de mogelijke antwoorden, dan zullen veel respondenten dit antwoord kiezen om te voorkomen dat ze na moeten denken. Weglaten van “Weet niet” dwingt respondenten echter tot een mening die ze misschien niet hebben. Dan vluchten ze vaak naar een neutrale middencategorie (‘niet tevreden en niet ontevreden’).

Een mogelijke oplossing is de toevoeging van een *filtervraag*. Daarin vraagt u de respondenten eerst of ze een mening hebben over een bepaalde kwestie. En alleen als ze een mening hebben, komt u met een vervolgvraag waarin de respondent die mening nader moet specificeren door het kiezen van de overeenkomstige antwoordcategorie. In hoofdstuk 6, paragraaf 6.3 (Peilingen via het internet), gaan we nu wat dieper in op het probleem van “Weet niet”.

Als u een reeks gesloten vragen stelt die allemaal dezelfde antwoordmogelijkheden hebben, dan zou u kunnen overwegen ze te combineren tot een *matrixvraag*. Elke rij in de matrix correspondeert dan met een vraag en elke kolom met een mogelijk antwoord. Hier is een voorbeeld:

	Uitstekend	Heel goed	Goed	Redelijk	Slecht
Wat vindt u in het algemeen van de kwaliteit van de omroep?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wat vindt u van de kwaliteit van het nieuwsprogramma's?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wat vindt u van de kwaliteit van het sportprogramma's?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wat vindt u van de kwaliteit van het muziekprogramma's?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wat vindt u van de kwaliteit van het culturele programma's?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wat vindt u van de kwaliteit van het kinderprogramma's?	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Op het eerste gezicht lijkt een matrixvraag wel voordelen te hebben. Een dergelijke vraag neemt minder ruimte in beslag dan een reeks losse vragen. En hij biedt de respondenten ook wat meer overzicht. Daarom zou het beantwoorden van de vragen wel eens minder tijd kunnen kosten.

Bij een schriftelijke enquête of een online-enquête vullen de respondenten zelf de vragenlijst in. Het beantwoorden van een reeks vragen in een matrix vereist wel enige inspanning. Het is cognitief ingewikkeld. De respondenten kunnen op allerlei manieren de matrix doorlopen. Daarom is het risico groot dat ze vragen overslaan.

Als de matrixvraag een onderdeel is van online-enquête, dan is het mogelijk dat de matrix op het scherm van de respondent maar voor een deel zichtbaar is. Dat hangt af van de omvang en resolutie van het beeldscherm. Als de matrix niet volledig zichtbaar is, dan kun de respondenten vragen of antwoordmogelijkheden al snel over het hoofd zien.

Bij het invullen van vragenlijsten kunt u te maken krijgen met allerlei vormen van *satisficing*. Die term is een samenvoeging van de woorden *satisfy* en *suffice*. Satisficing komt erop neer dat respondenten niet meer hun uiterste best doet om het enig juiste antwoord op de vraag te geven. In plaats daarvan kiezen ze snel een antwoord dat er wel redelijk acceptabel uitziet. U krijgt dus niet een juist antwoord, maar een makkelijk antwoord.

Bij matrixvragen bestaat het risico van satisficing, vooral wanneer een enquête een hele reeks saaie matrixvragen bevat. Satisficing kan de vorm aannemen van *straight-lining*. Respondenten verliezen hun motivatie en kruisen uit gemakzucht alle antwoorden in dezelfde kolom aan. Lekker snel en makkelijk! En bij voorkeur kiezen ze dan ook nog de veilige neutrale middencategorie (zoals in het voorbeeld).

Het kan helpen om enige afwisseling aan de brengen in de manier waarop je de vragen moet beantwoorden. Om je gedrag te beschrijven moet je dan bij de ene vraag ‘mee eens’ antwoorden en bij de andere vraag juist ‘mee oneens’. Die afwisseling zou de respondenten beter bij de les kunnen houden en bovendien kunt u zo ook ontdekken of er bij sommige respondenten sprake is van satisficing.

Vanwege de hierboven genoemde problemen is het verstandig om het gebruik van matrixvragen zoveel mogelijk te beperken. Als het dan toch moet, doe het dan zo goed mogelijk. Wat de respondent bijvoorbeeld helpt, is het geven van een afwisselende achtergrondkleur aan de rijen.

Gesloten vraag, meer antwoorden

Meestal moet de respondent bij een gesloten vraag maar één antwoord uit de lijst kiezen. Soms komt het echter voor dat meer dan één antwoord in aanmerking komt. Dan kunt u de respondenten de mogelijkheid bieden om meer antwoorden aan te kruisen:

Over welke aspecten van het verkeer in uw beurt bent u ontevreden?
(Meer antwoorden zijn mogelijk)

- De parkeergelegenheid
- De afstand tot de bushalte
- De busverbindingen
- De genomen verkeersmaatregelen
- De straatverlichting
- De veiligheid voor fietsers
- De veiligheid voor voetgangers
- Andere aspecten

Het kan heel goed zo zijn dat iemand ontevreden is over meer dan één aspect van het verkeer. Dan is het verstandiger meer dan één antwoord toe te staan. Merk ook op dat dit voorbeeld zich indekt tegen de situatie dat iemand met een antwoord komt dat niet in de lijst staat. Daarvoor is de categorie ‘Andere aspecten’ opgenomen.

Bij dit soort gesloten vragen kan ook *satisficing* optreden, vooral als de lijst met mogelijke antwoorden lang is. De respondent kruist vooraan in de lijst een paar antwoorden aan en houdt het dan verder voor gezien. In feite is er ook nog sprake van een *primacy effect*, want vooral antwoorden vooraan in de lijst krijgen de voorkeur.

Numerieke vraag

Een ander vraagtype dat veel voorkomt, is de *numerieke vraag*. U kunt daarbij denken aan vragen over prijzen, inkomen of leeftijden. Een veel voorkomende vraag in een peiling is:

Uit hoeveel personen bestaat het huishouden? _ _

Met een aantal vakjes of lijntjes kunt u aangegeven uit hoeveel cijfers het antwoord mag bestaan. Dit geeft wat extra aansporing tot het geven van een juist antwoord.

Merk op dat bij veel numerieke vragen het voor de respondent niet makkelijk is om het exacte antwoord te geven, omdat hij dat exacte antwoord niet weet en het veel werk is om het op te zoeken of te bepalen. Voorbeelden daarvan zijn vragen naar het netto jaarinkomen en het aantal uren en minuten dat iemand gisteren op het internet bezig was. Om dit soort problemen te vermijden, kunt u overwegen om van de numerieke vraag een gesloten vraag te maken:

Hoe lang was u gisteren actief op het internet?

- Helemaal niet
- 1 uur of minder
- 1 tot 2 uur
- 2 tot 3 uur
- 3 tot 4 uur
- 4 uur of langer

4.3 De volgorde van de vragen

Zijn alle vragen op een correcte manier geformuleerd, dan moet u nadenken over de volgorde waarin u de vragen gaat stellen. Het is in ieder geval goed om vragen over hetzelfde onderwerp te groeperen. Verdere verdient het aanbeveling om de vragenlijst met een paar interessante vragen te beginnen. Bewaar saaie, oninteressante vragen (bijvoorbeeld over opleiding over inkomen) maar voor het laatste deel van de vragenlijst. Zo houdt u de aandacht van de respondent beter vast.

Verder moet u bedenken dat je het antwoord op een vraag kunt beïnvloeden door vragen die u daarvoor stelt. Die eerdere vragen kunnen als het ware de toon zetten voor de respondent.

Voorbeeld 4.1. Beïnvloeding van de respondent

Will Tiemeijer vergelijkt in zijn boek “Wat 93,7 procent van de Nederlanders moet weten over opiniepeilingen” twee verschillende peilingen die op het zelfde moment in 2007 zijn uitgevoerd en die beide vragen naar de mening over Europese Unie.

De ene peiling was van Maurice de Hond. Daaruit bleek dat 43% van de Nederlanders de voordelen van de Europese Unie groter vond dan de nadelen, terwijl 35% de nadelen groter vond dan de voordelen. De rest wist het niet. Een niet zo erg positief beeld dus.

De andere peiling was de *Eurobarometer*. Dit is een peiling die in opdracht van de Europese Commissie wordt uitgevoerd. Daaruit bleek dat 69% van de Nederlanders het lidmaatschap van de EU een voordeel vond. En 77% vond het een goede zaak dat Nederland lid is van de EU. Dit is een veel positiever beeld.

Bij nader onderzoek blijkt dat in de peiling van Maurice de Hond de mensen in de voorafgaande vraag moesten aangeven wat ze de grootste nadelen van de EU vonden. Daarbij konden ze kiezen uit antwoorden als de snelle uitbreiding, de mogelijke toetreding van Turkije, de invoering van de euro, het bemoeien met zaken die beter door de landen zelf geregeld kunnen worden, de verspilling van geld in Brussel, het tekort aan democratische rechten voor de burgers, het verlies aan eigen identiteit, enz. Kortom, respondenten werden heel erg bewust gemaakt van de mogelijk negatieve aspecten van de EU. Positieve aspecten kwamen niet aan de orde. Dat had gevolgen voor de beantwoording van de volgende vraag. De Eurobarometer bevatte geen vragen met een dergelijke negatieve lading.

U dient te bedenken dat een vraag vooraan in het interview een heel ander effect kan hebben dan een vraag achteraan. Dit verschijnsel is bijvoorbeeld aangetoond in onderzoek naar woningbehoeften onder mensen. Men kon eerst heel tevreden zijn over de eigen woning, maar als na allerlei vragen (heeft u een ligbad? Heeft u een tuin op het zuiden? Heeft u wel eens lekkage gehad?) bleek dat er toch nog wel het een en ander ontbrak, dan kon tevredenheid wel eens omslaan in ontevredenheid. In

het algemeen kunnen we stellen dat een enquête de respondenten veel meer bewust kan maken van de problematiek die aan de orde komt. Daardoor worden ze aan het denken gezet, en dat kan hun mening wijzigen.

Een tweede aspect van de volgorde van de vragen is dat niet elke vraag relevant hoeft te zijn voor elk respondent. In een peiling over werk en werkloosheid zult u mensen met en zonder werk ondervragen. Werkenden vraagt u dan niet hoe ze zoeken naar werk, en werklozen vraagt u niet naar hun werkomstandigheden. Irrelevante vragen kunnen leiden tot irritatie bij de respondent en uiteindelijk zelfs tot een weigering om verder mee te werken. Bovendien zal de het voor een respondent niet eenvoudig zijn om irrelevante vragen te beantwoorden. Om deze problemen te vermijden, moet u, waar nodig, *sprongopdrachten* opnemen in de vragenlijst. Dat zijn instructies die aangeven dat een respondent bepaalde vragen moet overslaan, en de beantwoording verderop in de vragenlijst moet voortzetten.

Voorbeeld 4.2. Luisteronderzoek

1. Kent u Lommerdal FM, de lokale omroep van de gemeente Lommerdal?

- Nee → Vraag 5
- Ja → Vraag 2

2. Luistert u wel eens naar Lommerdal FM?

- Nee → Vraag 3
- Ja → Vraag 4

3. Waarom luistert u nooit naar Lommerdal FM?

- Luistert nooit naar de radio
- Geen belangstelling voor lokale gebeurtenissen
- Kan de lokale omroep niet goed ontvangen
- Andere reden

} → Klaar met invullen

4. Naar welk programma's van de lokale omroep luistert u regelmatig?
(meer antwoorden zijn mogelijk)

- Programma's met lokaal nieuws en informatie
- Sportprogramma's
- Programma's over kunst en cultuur
- Kerkelijke programma's
- Muziekprogramma's
- Andere programma's

5. Wat is uw geslacht?

- Man
- Vrouw

6. Wat is uw leeftijd?

- 13-19 jaar
- 20-39 jaar
- 40-59 jaar
- 60 jaar en ouder

Voorbeeld 4.2 laat zien hoe u sprongopdrachten in een vragenlijst kunt opnemen. Het is een verkorte versie van een denkbeeldig luisteronderzoek van een lokale omroep.

Als respondent de lokale omroep niet kennen, dan hoeven ze ook geen vragen over de omroep te beantwoorden. Daarom springen deze respondenten naar de algemene vragen aan het einde van de vragenlijst. Hetzelfde geldt voor respondenten die de omroep wel kennen, maar er niet naar luisteren. Deze respondenten moeten echter eerst wel uitleggen waarom ze niet luisteren.

Sprongopdrachten hebben nog het extra voordeel dat ze ook het aantal vragen beperken dat de respondenten moeten beantwoorden. Daardoor kan het interview sneller worden afgerond. Wel is het zo dat het ontwikkelen van een vragenlijst met sprongopdrachten ingewikkelder kan zijn. U moet zorgvuldig controleren of elke respondent wel het juiste pad door de vragenlijst neemt.

4.4 Testen van vragenlijsten

U moet een vragenlijst altijd eerst goed testen voordat hij het veld in gaat. Elke onderzoeker zal het met deze aanbeveling eens zijn, maar dat betekent niet dat het in de praktijk ook altijd gebeurt. Het komt regelmatig voor dat er niet genoeg tijd en geld is om een goede test uit te voeren. Daarom wordt aan dit aspect wel eens te weinig aandacht besteed.

Wat houdt een testprocedure precies in? Concreet betekent het dat u de vragenlijst in de praktijk uitprobeert. Dan kan op twee manieren gebeuren. In de eerste plaats kunnen de enquêteurs de respondenten benaderen en de echte interviewsituatie naspelen. De respondenten weten dan dus niet dat het slechts om een test gaat. Dat heeft het voordeel dat ze zich net zo gedragen als in een echte interviewsituatie. Ze zullen zich serieus gedragen en niet van hun normale gedrag afwijken omdat het toch maar om een test gaat. Een tweede manier om de test aan te pakken is de respondenten vertellen dat het om een test gaat. Dat heeft het voordeel dat de enquêteurs door kunnen vragen over hoe de respondenten de vragen hebben begrepen, wat precies onduidelijk was in een vraag, en waarom ze een bepaald antwoord hebben gegeven. We noemen dit ook wel *cognitief interviewen*.

Wat moet en kun je testen aan een vragenlijst? Een van de belangrijkste aspecten is de *validiteit*. De vragen, en dus de vragenlijst, moet op correcte wijze meten wat u beoogt te meten. Het is geen eenvoudige zaak om de validiteit vast te stellen. In de praktijk kunt u wel een aantal aspecten onderzoeken die allemaal op een of andere manier te maken hebben met validiteit.

In de eerste plaats moet het voor de respondenten duidelijk zijn wat u met de vraag bedoelt. De respondenten moeten een vraag op dezelfde manier interpreteren als u dat zelf doet. De literatuur over vragenlijsten staat boordevol voorbeelden van grote en kleine misverstanden. Zo vroeg een Amerikaanse onderzoeker naar 'heavy traffic in the neighborhood' en doelde daarmee op overlast van vrachtwagens. Veel respondenten dachten echter dat het ging om overlast veroorzaakt door drug dealers.

In een ander onderzoek werd een vraag gesteld over ‘family planning’. De onderzoeker doelde hiermee op het aantal kinderen dat het gezin wilde hebben, maar sommige ondervraagden dachten dat het ging om sparen van geld voor de komende vakantie.

Het nagaan of de respondenten de vragen op de juiste manier begrijpen, is een van de belangrijkste aspecten van het testen van de vragenlijst. Onderzoek heeft uitgewezen dat de respondent de vraag vaak op een heel andere manier begrijpt dan de onderzoeker bedoelt. Een respondent die niet goed weet wat hij moet antwoorden, past de vraag in zijn hoofd aan op zo’n manier hij hem wel kan beantwoorden.

Een tweede aspect is of een vraag voldoende variatie in de antwoordmogelijkheden biedt. Een vraag waarop bijna iedereen hetzelfde antwoord geeft, levert weinig informatie op. In het algemeen moet er zoveel variatie in antwoordmogelijkheden zijn dat je op basis van de antwoorden allerlei relevante deelgroepen in de populatie kunt onderscheiden. Er zijn uitzonderingen waarin een heel scheve antwoordersverdeling, bijvoorbeeld 99% ja en 1% nee, wel relevant is. Denk bijvoorbeeld aan het bepalen van het percentage analfabetisme, of het opsporen van een kleine groep die het risico loopt op een bepaalde ernstige ziekte.

Al is de betekenis van een vraag duidelijk, dan nog kan het lastig zijn om hem te beantwoorden. Een vraag als

Hoeveel minuten was u vorige week actief op het internet? _ _ _

is heel duidelijk en begrijpelijk, maar toch niet te beantwoorden, simpelweg omdat de respondent die informatie niet heeft, of alleen met zeer veel moeite kan achterhalen. Een bijna even lastige vraag is die naar het netto jaarinkomen van een gezin. Als u dit soort vragen toch wilt stellen, dient u zich te realiseren dat u hooguit een benadering van het goede antwoord krijgt. U zult daarmee genoeg moeten nemen.

Onderzoekers willen nog wel eens vergeten dat niet elke respondent evenveel enthousiasme kan opbrengen voor de vragen in de enquête. Naarmate de belangstelling van de respondenten voor de vragen en het onderwerpen afneemt, zal dit ook consequenties hebben voor de kwaliteit van de gegeven antwoorden. De enquêteurs dienen daarom in een test goed te registreren welke vragen interesse oproepen en welke vragen leiden tot desinteresse. Het zal duidelijk dat zijn dat een vragenlijst vol oninteressante vragen niet erg succesvol zal zijn.

Ook al zijn de vragen redelijk interessant, toch kan er tegen het einde van het interview een soort *vragenmoeheid* optreden. Dat zal natuurlijk speciaal bij lange vragenlijsten het geval zijn. Uitgangspunt moet altijd zijn de vragenlijst zo kort mogelijk te houden.

Het is niet alleen belangrijk om alle individuele vragen goed te testen. U moet ook naar de vragenlijst als geheel kijken. Zo moet de route door de vragenlijst op als natuurlijk overkomen bij de respondenten. Er moet als het ware een vanzelf-

sprekende opvolging van onderwerpen zijn. Een goed middel om dit na te gaan is de vragen hardop te laten voorlezen (in plaats van de vragenlijst zelf door te lezen). Al luisterend naar het verhaal vallen vaak de onnatuurlijke overgangen en wendingen beter op.

Al eerder is de suggestie gedaan om de vragenlijst met interessante vragen te beginnen. Saai, oninteressante achtergrondvragen (naar leeftijd, geslacht, inkomen, en zo) kunt u maar beter tot het einde bewaren. Daarmee voorkomt u potentiële problemen aan het begin van de vragenlijst.

Overigens is het soms toch nodig om achtergrondvragen aan het begin te stellen. Die situatie doet zich voor als een dergelijke vraag de rol van filtervraag speelt. Het antwoord op een *filtervraag* bepaalt de route door de vragenlijst. Als u mannen andere vragen wilt stellen dan vrouwen, dan zult u toch eerst naar het geslacht moeten vragen.

Vragenlijsten kunnen ingewikkelde sprongopdrachten bevatten. U moet controleren of al die route-instructies er wel toe leiden dat de respondenten in de voor hen relevante delen van de vragenlijst terechtkomen. Verkeerde sprongopdrachten kunnen ertoe leiden dat bepaalde vragen nooit worden gesteld (en dus ook nooit worden beantwoord) of dat de respondenten irrelevante vragen krijgen voorgeschoteld. Dit laatste kan leiden tot irritatie en zelfs tot weigering van verdere medewerking.

Ingewikkelde sprongopdrachten kunnen leiden tot veel verschillende paden door de vragenlijst. Het kan dan een heel werk zijn om elke mogelijk pad te controleren. Toch moet dat gebeuren, al is het maar om te voorkomen dat u bepaalde selectieve deelpopulaties verkeerd afhandelt in de vragenlijst.

Het is ook zinvol om te na te gaan hoe lang het gemiddeld duurt om een vragenlijst in te vullen. Het algemene advies is om de vragenlijst zo kort mogelijk te houden. Een wel gehanteerde regel is dat na zo ongeveer 50 minuten de aandacht van de respondent verslapt bij mondeling interviewen. De aanbeveling is dus het interview niet langer te laten duren (net zoals bij lesuren op school). Bij schriftelijk enquêteren en online enquêteren moeten de respondenten het helemaal zelf doen. Ze houden het dan meestal geen 50 minuten vol. Vragenlijsten moeten dan veel korter zijn. Denk hierbij aan, pakweg, 15 minuten.

De interviewers moeten er tijdens de test proberen achter te komen of het onderwerp van de enquête de belangstelling van de respondenten wekt. Als blijkt dat de vragenlijst te lang en te saai wordt gevonden, is het wellicht beter om de deze te heroverwegen. Een mogelijke remedie is het verminderen van het aantal vragen, maar daar zal geen enkele onderzoeker enthousiast over zijn. Een andere oplossing zou kunnen zijn wat meer variatie aan te brengen door het toevoegen van vragen over andere onderwerpen.

Een laatste aspect dat nog kan worden genoemd bij het testen van de vragenlijst is het algemeen welzijn van de respondenten. Tegenwoordig zien we enquêtes over een breed scala aan onderwerpen. Die ontzien vaak niet de persoonlijke levenssfeer

van de respondenten. Er zijn vragen over alcohol- en drugsgebruik, homoseksuele en heteroseksuele relaties, tevredenheid met het huwelijk, scheiding, mishandeling, dood van kinderen, eenzaamheid, geestelijke problemen, depressies, zelfmoord, lichamelijke handicaps, ongeneeslijke ziektes, religieuze ervaringen, angsten en geloof. De litanie is lang. Ondanks dat u altijd eerst om de vrijwillige medewerking van de respondent vraagt (het principe van de *informed consent*), moet u zich als onderzoeker toch altijd afvragen of vragen over dit soort onderwerpen geen nadelige invloed op de respondenten kunnen hebben. Na afloop van het interview moet de respondent zich niet ongelukkiger voelen dan ervoor.

Een goede test van een vragenlijst verloopt in twee fasen. In de eerste fase worden zo'n 25 tot 75 interviews afgenomen. De nadruk ligt vooral op het testen van de gesloten vragen. De bij elke vraag gegeven mogelijke antwoorden moeten duidelijk zijn. Elke respondent moet in staat zijn de voor hem relevante antwoord te vinden. En waar iemand niet het juiste antwoord kan vinden, moet er een 'nooduitgang' zijn in de vorm van de mogelijkheid 'Anders, nl ...'.

Het is een goed idee om de ervaringen van de interviewers in deze eerste testfase vast te leggen met een klein vragenlijstje. Vragen die zouden kunnen worden gesteld, zijn bijvoorbeeld:

- Waren er vragen waarbij de respondent zich ongemakkelijk voelde?
- Waren er vragen die moesten worden herhaald?
- Waren er vragen die de respondent verkeerd interpreteerde?
- Welke vragen waren het moeilijkst of vervelendst om te stellen?
- Waren er onderdelen in de vragenlijst waarin het interview zich voortsleepte?
- Waren er onderdelen in de vragenlijst waarbij het gevoel opkwam dat de respondent meer had willen zeggen?

De eerste fase moet een grondig onderzoek van de vragenlijst zijn dat de essentiële fouten in de vragenlijst aan het licht brengt. De tweede fase van de test is een soort generale repetitie. Daarin gaat het niet meer om het repareren van grote fouten, of het proberen van een geheel nieuwe aanpak. Het gaat om de 'fine tuning' en 'finishing touch', het wegwerken van de laatste loshangende rafeltjes. In de tweede fase simuleert u het echte interview. U vertelt de respondenten niet dat het om een test gaat. In deze fase zou u eventueel de mening van externe deskundigen over de vragenlijst kunnen vragen. De respondenten moeten 'echte' respondenten zijn. Denk ook hier weer aan een groep van 25 tot 75 mensen.

Grote onderzoeksorganisaties hebben voor het testen van vragenlijsten vaak een vragenlaboratorium ingericht. In een huiskamersituatie wordt op video vastgelegd wat er gebeurt tijdens het invullen van de vragenlijst. Eventueel kunnen onderzoekers door eenzijdige spiegels meekijken. Ook al is een dergelijke professionele testomgeving niet beschikbaar, dan is het nog steeds erg zinvol om de vragenlijst te testen op een beperkt aantal willekeurige personen. Problemen zullen dan al snel aan het licht komen.

5. Het trekken van een steekproef

Hoe kunt u op verantwoorde wijze een steekproef trekken uit een populatie? U zou natuurlijk zomaar wat objecten uit de losse pols kunnen aanwijzen. Een mooi voorbeeld hiervan is de methode die actualiteitenrubrieken op de TV soms hanteren. Voor de mening van de ‘Nederlander’ over een bepaald onderwerp sturen ze een verslaggever met een cameraman naar de Albert Cuypmarkt in Amsterdam. De verslaggever stelt vragen aan mensen op de markt. Het resultaat: een hoop leuke antwoorden, en wat je niet bevalt laat je weg. *Vox populi* (‘de stem van het volk’) noemen we dat wel. Deze aanpak levert een beeld op dat meestal geen goede afspiegeling is van de doelpopulatie. Daarom benadrukt een omroep als de BBC dat u *vox pops* hooguit ter illustratie mag gebruiken en nooit als een methode van onderzoek.

Hoe trekt u een steekproef die wel een goede afspiegeling is van de doelpopulatie? Daarvoor zijn twee ingrediënten nodig: een steekproefkader en een procedure voor het trekken van een steekproef uit dat steekproefkader.

5.1 Het steekproefkader

U wilt een steekproef trekken. Daarvoor hebt u een lijst nodig met namen, adressen of andere contactgegevens van ieder object in de doelpopulatie. Die lijst noemen we het *steekproefkader*.

Een *steekproefkader* is dus soort een administratieve weergave van de te onderzoeken populatie. Het is een lijst waarin elk object uit de doelpopulatie staat vermeld. Ook zal duidelijk moeten zijn hoe u contact kunt opnemen met die objecten ten behoeve van uw onderzoek. Bestaat de populatie uit mensen, dan kan het steekproefkader een lijst met adressen zijn. Denk aan een kaartenbak of een computerbestand. Een voorbeeld hiervan is een ledenadministratie. Er zijn natuurlijk ook andere manieren om met mensen in contact te komen, zoals telefoonnummers of e-mailadressen.

Vroeger, toen veel steekproefkaders nog bestonden uit kaartenbakken (bijvoorbeeld een bevolkingsregister van een gemeente of een ledenadministratie van een organisatie), moest je steekproeven daaruit met de hand trekken. Tegenwoordig zijn veel steekproefkaders beschikbaar in de vorm van computerbestanden. Dat maakt het trekken van een steekproef een stuk eenvoudiger en een stuk sneller.

Voorbeeld 5.1. Postcodes

Er zijn bedrijven in Nederland die postcodegegevens verkopen. Die bedrijven leveren bestanden waarin postcodes gekoppeld zijn aan straatnamen en huisnummers. Er is dan in feite sprake van een adressenbestand. U kunt dat bestand gebruiken voor het trekken van steekproeven van adressen (en dus van huishoudens die op de adressen wonen).

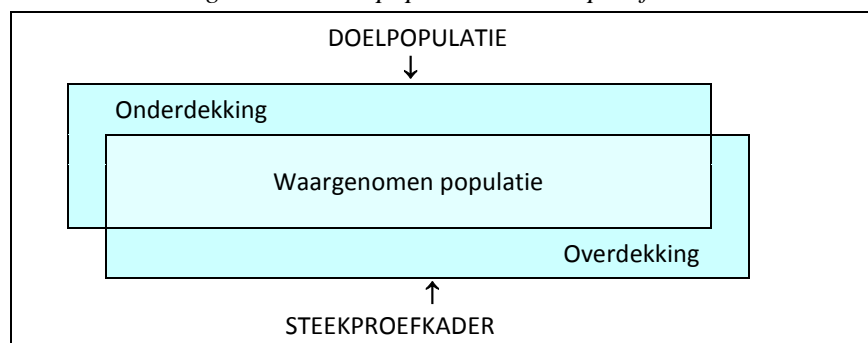
De tabel hieronder bevat een fragment van een dergelijk bestand. Het betreft adressen in de kern Hazerswoude-Rijndijk van de gemeente Rijnwoude.

Postcode	Straat	Reeks	Woonplaats
2394 BK	Da Costasingel	26 t/m 32	HAZERSWOUDE-RIJNDIJK
2394 BL	Da Costasingel	33 t/m 39	HAZERSWOUDE-RIJNDIJK
2394 BL	Da Costasingel	34 t/m 42	HAZERSWOUDE-RIJNDIJK
2394 BM	Bruggestraat	1 t/m 5	HAZERSWOUDE-RIJNDIJK
2394 BM	Bruggestraat	2	HAZERSWOUDE-RIJNDIJK
2394 BZ	Groenendijksepad	1 t/m 5	HAZERSWOUDE-RIJNDIJK
2394 BZ	Groenendijksepad	2 t/m 8	HAZERSWOUDE-RIJNDIJK
2394 CA	Rijndijk	147 t/m 167	HAZERSWOUDE-RIJNDIJK
2394 CB	Rijndijk	169 t/m 197	HAZERSWOUDE-RIJNDIJK
2394 CC	Rijndijk	199 t/m 225	HAZERSWOUDE-RIJNDIJK

Bestanden als dit zijn verkrijgbaar voor heel Nederland, maar ook voor kleinere regio's, zoals gemeenten.

Het is van groot belang dat het steekproefkader overeenkomt met de doelpopulatie. Is dat niet het geval dan doet u in feite uitspraken over een andere populatie dan uw doelpopulatie. Wat er gebeurt als u, door praktische problemen gedwongen, een steekproef trekt uit een steekproefkader dat niet overeenkomt met de doelpopulatie, is schematisch weergegeven in figuur 5.1.

Figuur 5.1. Doelpopulatie en steekproefkader



In de eerste plaats kunnen er objecten in de doelpopulatie zitten die niet zijn opgenomen in het steekproefkader. Deze situatie heet *onderdekking*. Dit verschijnsel doet zich bijvoorbeeld voor als u onderzoek wilt doen onder illegaal hier verblijvende mensen terwijl u een steekproef hebt getrokken uit het bevolkingsregister.

De gevolgen van onderdekking kunnen ernstig zijn. Als een selecte groep uit de doelpopulatie daardoor ontbreekt in uw onderzoek, kunnen de resultaten een vertekend beeld opleveren. Daarbij komt bovendien dat onderdekking niet eenvoudig valt waar te nemen als u er niet echt goed naar zoekt.

Het kan ook gebeuren dat er objecten in het steekproefkader zitten die niet thuishoren in de doelpopulatie. Dat verschijnsel duiden we aan met *overdekking*. Het ten onrechte meenemen van deze objecten in het onderzoek kan ook tot verkeerde conclusies leiden. Overdekking kunt u vrij eenvoudig ontdekken in het veld, door voor elk object in de steekproef eerst te controleren of het wel voldoet aan de definitie van de doelpopulatie.

Voorbeeld 5.2. Onderdekking en overdekking

Stel dat u een telefonische peiling wilt doen onder de gezinnen in een gemeente. Daarvoor zou u het telefoonboek als steekproefkader kunnen gebruiken.

In de eerste plaats kan er dan onderdekking optreden als gevolg van het verschijnsel van onvermelde nummers. Pakweg 30% tot 40% van de mensen staat tegenwoordig niet meer in het telefoonboek. Daarbij zitten veel mensen die alleen maar een mobieltje hebben. Het gaat hier vooral om jongeren. Dus die groep zou wel eens zwaar ondervertegenwoordigd kunnen zijn in uw onderzoek.

In de tweede plaats staan er nummers in het telefoonboek die niet bij gezinnen horen. Denk hierbij aan nummers van winkels en bedrijven. Daardoor kan het gebeuren dat u mensen belt en enquêteert die niet in de populatie thuishoren. Dit is overdekking.

Problemen kunnen ook ontstaan als doelpopulatie en steekproefkader niet dezelfde eenheden bevatten. Een bekend voorbeeld daarvan is de situatie waarin u een steekproef van personen wilt trekken uit een steekproefkader met adressen. U zou dan eerste een steekproef van adressen kunnen trekken, en vervolgens op elk geselecteerd adres een willekeurige persoon kunnen kiezen. U maakt dan een essentiële denkfout als u doet alsof iedereen dezelfde kans heeft gehad om in de steekproef te komen. Immers, personen in grote gezinnen hebben een veel kleinere kans dan personen in kleinere gezinnen. De trekkingskans is omgekeerd evenredig met de omvang van het gezin. We komen hierop verder in dit hoofdstuk nog terug.

Voorbeeld 5.3. Een peiling in een gemeente

Stel dat u de mening wilt peilen van de inwoners in een gemeente. Wat voor steekproefkader zou u dan hiervoor kunnen gebruiken? Er zijn een paar mogelijkheden. We gaan hier in op het gebruik van het bevolkingsregister, een adressenbestand, een postcodebestand, het telefoonboek en Random Digit Dialing (RDD):

- Het *bevolkingsregister*. Dit wordt ook wel de Gemeentelijke Basis Administratie (GBA) genoemd. Het bevat alle mensen die (legaal) in de gemeente verblijven. Een steekproef hieruit levert een keurige lijst met namen en adressen. Er zijn geen problemen met overdekking en onderdekking, tenzij u ook illegaal in de gemeente wonende personen in uw onderzoek wilt meenemen. Lastig is dat een gemeente alleen steekproeven mag trekken uit het GBA voor wetenschappelijke onderzoeksinstituten en het onderzoek een algemeen belang dient. Dit betekent dat de gemeente verzoeken voor het trekken van een steekproef vaak zal weigeren.
- Er zijn bedrijven die (tegen betaling) een *adressenbestand* kunnen leveren. Die bestanden zijn meestal gebaseerd op het postafgiftepuntenbestand van PostNL. U kunt daarbij eventueel selectiecriteria opgeven. Bijvoorbeeld alleen personen uit een bepaalde regio, leeftijdsklasse, welstandsklasse of type woning. U moet daarna nog wel per adres een of meer personen selecteren.
- Er zijn bedrijven die een *postcodebestand* kunnen leveren. Elke regel van dit bestand bevat een postcode (4 cijfers + 2 letters) en de adressen die daarbij horen. Uit zo'n bestand kunt u een adressenbestand maken. U kunt echter niet zien of het om woonhuizen of bedrijven gaat. Er zullen daarom veel onbruikbare adressen tussen zitten.
- Het *telefoonboek*. U kunt een steekproef trekken uit het telefoonboek voor zover dit betrekking heeft op de desbetreffende gemeente. Al eerder is gemeld dat dit steekproefkader verre van goed is. Het lijdt aan behoorlijke onderdekking omdat veel mensen er niet in staan. En er is ook sprake van overdekking omdat het telefoonnummers van zowel huishoudens als bedrijven bevat.
- Onderdekking maakt dat voor een peiling het telefoonboek niet een ideaal steekproefkader is. U kunt ook nog op een andere manier aan telefoonnummers komen. Dat heet *Random Digit Dialing*. U laat de computer willekeurige telefoonnummers genereren. U kunt dat bijvoorbeeld doen door een telefoonnummer uit het telefoonboek te nemen en vervolgens het laatste cijfer te vervangen door een willekeurig ander cijfer. Nadeel is dat u vooraf totaal geen idee hebt wie u belt. Soms is er ook geen verschil te horen tussen niet bestaande nummers (die u verder kunt vergeten) en nummers die niet worden opgenomen (zodat u het later nog eens moet proberen). En u moet zich afvragen wat u moet doen met telefoonnummers die zijn opgenomen in het *Bel-me-niet Register*. Weliswaar verkoopt u niets, maar toch.

Het is in de praktijk lang niet altijd eenvoudig om een geschikt steekproefkader te vinden. Daarom zou u uw toevlucht kunnen nemen tot een andere manier om een steekproef te trekken uit een populatie die niet is gebaseerd op loting uit een steekproefkader. Daarmee zet u wel de validiteit van de uitkomsten van uw peiling

op het spel. Uw steekproef hoeft dan geen afspiegeling meer te zijn van de populatie waaruit hij is getrokken. Schattingen van populatiekenmerken kunnen de plank dus volledig misslaan.

Voorbeeld 5.4. Hoe het niet moet

We geven drie voorbeelden van het trekken van een steekproef die u niet in de praktijk moet toepassen, omdat ze gegarandeerd tot verkeerde uitkomsten leiden.

Peilen in het winkelcentrum

Een lokale omroep wilde een luisteronderzoek uitvoeren. Om toch maar snel een heleboel gegevens te verzamelen, besloot de onderzoeker om enquêteurs op zaterdagmiddag naar het plaatselijke winkelcentrum te sturen. Het was daar gezellig druk zodat veel mensen konden worden aangesproken. In korte tijd was er een al een hele stapel vragenlijsten ingevuld.

Bij de analyse van de gegevens kwamen de onderzoeker tot een onverwachte conclusie: bijna niemand luisterde naar het sportprogramma dat op zaterdagmiddag werd uitgezonden.

Bij nader inzien is die conclusie natuurlijk helemaal niet zo verrassend. Er werden alleen maar mensen ondervraagd die op zaterdagmiddag in het winkelcentrum aanwezig waren. Daarmee werd de doelpopulatie in feite gereduceerd van alle inwoners tot het zaterdagse winkelpubliek. En het is niet zo verbazingwekkend dat die op zaterdagmiddag niet naar de radio luisterden. De mensen in de steekproef waren geen goede afspiegeling van alle inwoners van de gemeente.

De peiling van de ANWB over het rekeningrijden

In de eerste maanden van 2010 was er in Nederland een uitgebreide discussie over het invoeren van rekeningrijden. De ANWB besloot de mening van de leden hierover te peilen. Op de website van de ANWB werd een uitgebreide vragenlijst gezet. Iedereen kon die vragenlijst invullen, en dus ook bijvoorbeeld buitenlanders. Het was zelfs mogelijk de vragenlijst meer dan één keer in te vullen.

Door deze opzet kun je vraagtekens zetten bij de representativiteit van de peiling. Om te beginnen kun je je afvragen wat eigenlijk de doelpopulatie is. De ANWB leden? Alle autorijders? Alle Nederlanders?

De ANWB gaf aan dat het om de leden ging. Gelukkig werd in de peiling de vraag gesteld of de respondent lid was van de ANWB. Het is echter nog maar de vraag of iedereen die zijn mening kwijt wilde, die vraag ook naar eer en geweten heeft beantwoord.

Uiteindelijk werd de vragenlijst ingevuld door 350.000 mensen die zeiden lid te zijn. Daarvan was 68% voorstander van een vorm van rekeningrijden. In hun onderzoeksrapport vermeldden de onderzoekers zelf ook dat de peiling niet representatief was. Ze gaven aan dat voor hen belangrijker was dat veel mensen meededen aan het onderzoek en hun mening gaven.

Het lezersonderzoek van een tijdschrift

De redactie van een gratis huis-aan-huisblad wilde weten of het blad goed werd gelezen. Daarom werd besloten tot het uitvoeren van een lezersonderzoek. In het blad werd een vragenlijst opgenomen. En uiteraard was een van de vragen of men het blad wel eens las. De respondenten moesten de ingevulde vragenlijst uitknippen en opsturen naar de redactie.

Tot de niet geringe verbazing en blijdschap van de redactie bleken bijna alle respondenten het blad wel eens te lezen. De redactie trok dan ook de conclusie dat het blad goed werd gelezen.

De adder die hier onder het gras zat, was dat mensen die het blad niet lasen, het onmiddellijk ongezien weggooiden, en dus ook de vragenlijst helemaal niet tegenkwamen. In feite werd de doelpopulatie beperkt tot alleen mensen die op zijn minst het blad wel eens doorbladerden. De redactie hield zichzelf voor de gek. Jammer, want zo werd het bijvoorbeeld ook niet duidelijk waarom al die andere mensen het huis-aan-huisblad niet lasen.

5.2 De steekproef

U kunt op aller manieren een steekproef trekken uit een populatie, maar er is slechts één manier om het goed te doen en dat is via een *kanssteekproef*. Ieder object moet een positieve kans hebben om in de steekproef terecht te komen en al die kansen moeten bekend zijn. Alleen dan hebben uw schattingen geen systematische afwijking. En alleen dan kunt u de onzekerheidsmarges van de uitkomsten uitrekenen. Die onzekerheidsmarges geven aan hoe ver uw schatting maximaal van de werkelijkheid kan afliggen. Kortom, als u met uw peiling uitspraken wilt doen over een populatie als geheel, dan zult u de steekproef door loting moeten trekken.

Op het eerste gezicht lijkt het voor de hand te liggen aan het lotingmechanisme de eis op te leggen dat elk object in de populatie *dezelfde* kans moet hebben om in de steekproef te komen. Inderdaad is dit een van de manieren waarop u een steekproef kunt trekken, en misschien is het ook wel een van de meest gebruikte manieren. Maar het is ook mogelijk om steekproeven met ongelijke kansen te trekken. Later komen we hier nog op terug.

Representatieve steekproef

De term *representatief* kom je vaak tegen als het over steekproeven gaat. Door te beweren dat een steekproef representatief is, wordt de suggestie gewekt dat het een

goede, betrouwbare steekproef is. Helaas staat er dan meestal niet bij wat ‘representatief’ precies betekent. De praktijk leert dat deze term in wisselende betekenissen en met wisselende bedoelingen wordt gebruikt. Soms is het niet meer dan een ‘vlag op de modderschuit’. De onderzoeker noemt de steekproef representatief en stelt zo de lezer of toeschouwer gerust. ‘Maak je maar geen zorgen, het zit wel goed met de steekproef’. Een ander, en een stuk beter, gebruik van de term representatief is om aan te geven dat de steekproef ‘eerlijk’ is getrokken. Dat wil zeggen dat geen enkele object in de doelpopulatie een voorkeursbehandeling heeft gehad. Soms ook betekent representatief dat bepaalde verhoudingen die in de doelpopulatie gelden, ook zijn terug te vinden in de steekproef. Zo betekent representatief met betrekking tot leeftijd dat de leeftijdsverdeling in de steekproef overeenkomt met die in de doelpopulatie. Het zal duidelijk zijn dat u maar beter niet over representatieve steekproeven kunt praten, tenzij u duidelijk uitlegt wat u daarmee bedoelt.

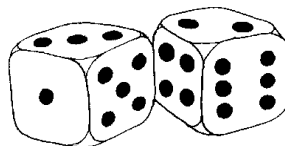
Loten met een aselector

U hebt dus een lotingsmechanisme nodig dat iedereen in de populatie een even grote kans geeft om in de steekproef te komen. In de statistiek noemen we zo'n eerlijke lotingmechanisme een *aselector*. Het is een ‘apparaat’ dat aan de volgende eisen voldoet:

- U kunt het apparaat steeds weer opnieuw gebruiken.
- Elke keer dat u het apparaat in werking stelt, produceert het één van de getallen 1 t/m N als uitkomst. We nemen aan dat de waarde van N bekend is.
- Elke keer opnieuw hebben alle N mogelijke uitkomsten dezelfde kans. Kennis over eerdere uitkomsten helpt u niet bij het beter voorspellen van de volgende uitkomst. Kortom, elk voorspellingssysteem faalt.

De aselector is een theoretisch concept. In de praktijk bestaat de ideale aselector niet. Er zijn wel ‘apparaten’ die in de buurt komen. Een simpel voorbeeld is een munt. Daarbij hebben ‘kop’ en ‘munt’ een even grote kans ($N = 2$). Een ander voorbeeld van een aselector is een dobbelsteen, waarbij de cijfers één t/m zes ($N = 6$) elk een even grote kans hebben (als de dobbelsteen tenminste ‘eerlijk’ is).

Figuur 5.2. Dobbelstenen



De munt is alleen te gebruiken als de populatie uit twee elementen bestaat, en de dobbelsteen in principe alleen voor populaties van zes elementen. Meestal wilt u een steekproef trekken uit een grotere populatie. Stel eens dat u een steekproef moet

trekken van 1.000 personen uit een gemeente van 18.000 inwoners. Er bestaan echter geen 18.000-zijdige dobbelstenen.

In de praktijk trekt u een steekproef met een rekenmachine of een computerprogramma. Die hebben vaak de mogelijkheid om willekeurige getallen te genereren uit het interval $[0, 1)$. Elke willekeurige waarde tussen 0 en 1 is mogelijk. De waarde 0 kan wel voorkomen, maar de waarde 1 net niet.

Voorbeeld 5.5. Aselecte getallen met een rekenmachine

De CASIO FX-82 rekenmachine heeft een knop RAN#. Elke keer dat u op die knop drukt, verschijnt er een willekeurige waarde uit het interval $[0, 1)$.

Het op een gegeven moment 20 keer drukken op deze knop leverde de volgende waarden op:

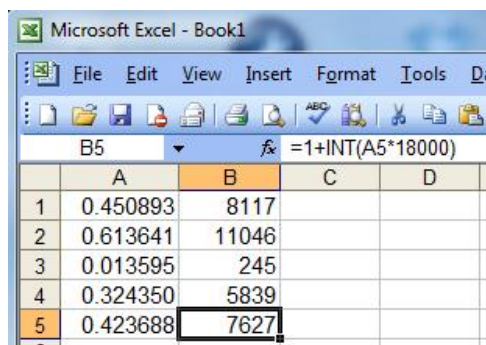
0,360	0,319	0,778	0,753	0,521	0,652	0,609	0,812
0,057	0,756	0,205	0,465	0,023	0,128	0,394	0,381
0,802	0,031	0,415	0,065				



Om een steekproef te trekken hebt u willekeurige getallen nodig uit de reeks $1/t/m N$, waarbij N de omvang van de populatie aanduidt. Dan hebt u dus niet onmiddellijk wat aan de willekeurige waarden uit het interval $[0, 1)$. U moet waarden omzetten in gehele getallen. Dat gaat als volgt:

- (1) Vermenigvuldig de willekeurige waarde met de omvang van de populatie. Dat levert waarden in het interval $[0, N)$. De waarde 0 kan wel voorkomen en de waarde N net niet.
- (2) Rond de verkregen waarde af naar beneden (door de cijfers achter de komma weg te laten). Dat levert een geheel getal op in de reeks van 0 tot $N-1$.
- (3) Tel 1 op bij het verkregen getal. Dat levert een geheel getal op in de reeks van 1 t/m N .

Figuur 5.3. Een steekproef trekken met een spreadsheet



U kunt ook een spreadsheet gebruiken voor het trekken van een steekproef. In bijvoorbeeld MS Excel kunt u dat snel en eenvoudig doen. Zie figuur 5.3 voor een

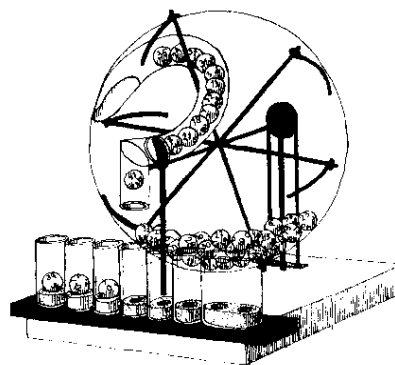
voorbeeld. Vul eerst een kolom met aselechte waarden tussen 0 en 1. Dat kan met de functie `ASELECT()` (Nederlandse versie) of `RAND()` (Engelse versie). Stel dat de omvang van de populatie 18.000 is. Genereer dan eerst willekeurige waarden in de cellen `A1, A2, ...`. Vervolgens kunt u de steekproefnummers in de cellen `B1, B2, ...` berekenen met de formules `=1+INTEGER(A1*18000)`, `=1+INTEGER(A2*18000)`, enz.

Steekproeven met en zonder teruglegging

Als u een paar keer met een dobbelsteen gooit, dan is het niet uitgesloten dat een bepaald aantal ogen meer dan één keer voorkomt. Datzelfde geldt ook voor het gebruik van de aselector. Als die een reeks getallen produceert, dan kan het best gebeuren dat een nummer verschillende keren voorkomt. De consequentie hiervan is dat de corresponderende persoon uit de populatie meer dan eens in de steekproef komt. Dit is niet zinvol, want het zou betekenen dat die personen de vragenlijst nogmaals moeten invullen. Dat levert geen nieuwe informatie op. Daarom gaat de voorkeur uit naar een *steekproef zonder teruglegging*. Daarin kan elk element hooguit één maal voorkomen.

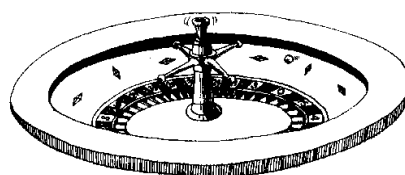
De lottomachine is een goed voorbeeld van een steekproef zonder teruglegging. Een geselecteerd nummer verdwijnt uit de populatie en daarom kunt u dit nummer niet nog een keer trekken.

Figuur 5.4. De lottomachine: een steekproef zonder teruglegging



De roulette is een goed voorbeeld van een *steekproef met teruglegging*. Bij elke beurt opnieuw hebben alle nummers (ook de al getrokken nummers) dezelfde kans (als de roulette tenminste eerlijk is). Het is dus heel goed mogelijk dat u een nummer nog een keer trekt.

Figuur 5.5. De roulette: een steekproef met teruglegging



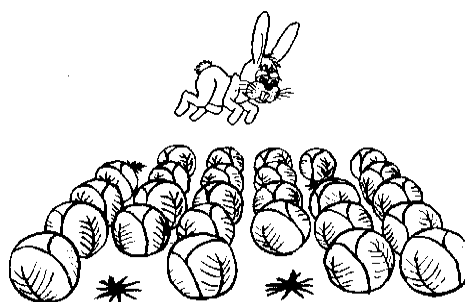
In de volgende paragrafen zullen we drie technieken voor het trekken van een steekproef wat nader bekijken:

- De *enkelvoudige aselecte steekproef*. U trek een steekproef met gelijke kansen en zonder teruglegging
- De *systematische steekproef*. Dat is een wat eenvoudiger manier om een enkelvoudige aselecte steekproef te trekken. Maar pas op: deze simplificatie werkt niet altijd goed.
- De *tweetrapssteekproef*. Dit is een aanpak waarin u eerst een steekproef van adressen trekt, en vervolgens op elk geselecteerd adres een persoon.

5.3 De enkelvoudige aselecte steekproef

De *enkelvoudige aselecte steekproef zonder teruglegging* is de eenvoudigste manier om een steekproef te trekken en misschien is het ook wel de bekendste manier om dat te doen. Deze manier van trekken staat het dichtst bij wat u intuïtief onder het trekken van een steekproef verstaat, namelijk dat elk object in de populatie dezelfde kans moet hebben om in de steekproef te komen. Het komt overeen met het vullen van een hoge hoed met lootjes. Voor elk persoon in de populatie is er een lootje. Vervolgens trekt u (na goed husselen) een aantal lootjes uit de hoge hoed.

Figuur 5.6. Een enkelvoudige aselecte steekproef zonder teruglegging



De omvang van de doelpopulatie geven we aan met N . Dan trekt u bij een enkelvoudige aselecte steekproef zonder teruglegging nummers uit de reeks van 1 t/m N . Elke nummer moet een even grote kans hebben. Een nummer mag hooguit maar één keer in de steekproef komen.

Wilt u een niet al te grote steekproef trekken, dan kunt u dit met een rekenmachine doen. Die rekenmachine moet dan wel een functie hebben die een willekeurige waarde tussen 0 en 1 genereert (waarbij 0 wel kan voorkomen en 1 net niet).

Een enkelvoudige aselecte steekproef trekken met een rekenmachine

Stap 1: Trek een willekeurige waarde RAN uit $[0, 1)$.

Stap 2: Vermenigvuldig die waarde met de omvang van de populatie N . Dit levert een waarde in het interval $[0, N-1)$.

Stap 3: Rond die waarde naar beneden af op een gehele waarde. Dit levert een getal uit de reeks van 1 t/m $N-1$.

Stap 4: Tel 1 op dit getal. Dit levert een volgnummer uit de reeks 1 t/m N .

Stap 5: Is dit getal al eerder in de steekproef getrokken, negeer het dan en doe een nieuwe poging. Ga hiervoor terug naar stap 1. Is het getal nog niet eerder getrokken, voeg het dan toe aan de steekproef.

Stap 6: Is de gewenste omvang van de steekproef nog niet bereikt, ga dan terug naar stap 1 en trek een nieuw object.

Zijn de populatie en de gewenste steekproef wat groter, dan kunt u overwegen de steekproef te trekken met een spreadsheet. Daar is minder handwerk voor nodig,

Een enkelvoudige aselechte steekproef trekken met Excel

Stap 1: Vul kolom A met de volgnummers van de objecten in de populatie.

Hiervoor kunt u de functie `ROW()` gebruiken. Die zet het corresponderende rijnummer in de cel. Vul evenveel cellen in de kolom als de populatie objecten heet.

Stap 2: Vul kolom B met willekeurige waarden uit het interval $[0, 1)$. Hiervoor kunt u de functie `ASELECT()` of `RAND()` gebruiken. Het spreadsheet-fragment linksonder bevat een voorbeeld.

Stap 3: Kies in het menu *Extra* voor *Opties* en zet in tabblad *Berekenen* de optie *Berekening op Handmatig*.

Stap 4: Selecteer kolom A en B, sorteer dit blok op kolom B. Het resultaat is zoiets als in het spreadsheet-fragment rechtsonder.

Stap 5: Uw steekproef bestaat nu uit de nummers in het bovenste deel van kolom A. Wilt u 10 objecten hebben, neem dan de eerste 10 nummers in deze kolom.

	A	B
1	1	0.677108
2	2	0.851208
3	3	0.902135
4	4	0.624257
5	5	0.914697
6	6	0.276631
7	7	0.880277
8	8	0.255902
9	9	0.543253
10	10	0.785882
11	11	0.862632
12	12	0.910895
13	13	0.490271

	A	B
1	903	0.000018
2	474	0.000964
3	286	0.001410
4	955	0.003571
5	982	0.003581
6	595	0.003699
7	399	0.004868
8	790	0.006207
9	689	0.006324
10	422	0.007093
11	292	0.007166
12	762	0.007327
13	510	0.007407

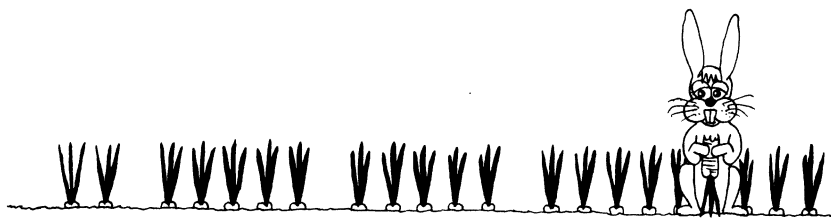
Voor grote steekproeven (van, zeg, een paar duizend objecten) uit een heel grote doelpopulatie is dit nog steeds een heel gedoe. Het is dan beter hiervoor een

computeralgoritme te ontwikkelen. De website www.survey-onderzoek.nl bevat een dergelijk algoritme. Daarmee kunt u een steekproef van maximaal 1.000 elementen trekken uit een populatie van willekeurige omvang.

5.4 De systematische steekproef

Als u een enkelvoudige aselechte steekproef echt met de hand trekt uit een lange lijst, dan kan dat behoorlijk veel werk zijn. Vooral als de elementen in die lijst niet zijn genummerd, is het niet eenvoudig om, bijvoorbeeld, element 348 eruit te halen. Voor dit soort situaties kan de *systematische steekproef* een alternatief bieden. Daarbij loopt u op een systematische wijze door het bestand. Het beginpunt bepaalt u door loting. Vervolgens slaat u steeds een vast aantal elementen over voordat u het volgende element selecteert. Uiteraard moet u de procedure wel zo inrichten dat elk object in de doelpopulatie dezelfde kans heeft om in de steekproef te komen.

Figuur 5.7. De systematische steekproef



Eerst moet u vaststellen hoeveel objecten er in de lijst staat, en hoeveel u daaruit wilt trekken. Met deze gegevens kunt u de *staplengte* bepalen. Dat is de lengte van de sprong waarmee u door de lijst springt. De staplengte krijgt u door de lengte van de lijst te delen door de omvang van de steekproef.

Voorbeeld 5.6. Berekening van de staplengte

De lijst met adressen voor een luisteronderzoek van een lokale omroep bevat 9.590 adressen. Als u daaruit een systematische steekproef van 500 adressen wilt trekken, dan is de *staplengte* gelijk aan 19,18. Immers, delen van 9.590 door 500 levert de uitkomst 19,18 op.

Vervolgens moet u het *startgetal* bepalen. Dat is het volgnummer van het object in de lijst waarmee u de steekproeftrekking begint. Het startgetal is een willekeurige waarde uit het interval dat loopt van 0 t/m de *staplengte*. Dit kunt u bijvoorbeeld bepalen met een handrekenmachine die een willekeurige waarde uit het interval [0, 1) kan genereren. Vermenigvuldig die waarde met de staplengte.

Als de waarden van de staplengte en het startgetal bekend zijn, dan ligt daarmee de steekproef vast. Het eerste object in de steekproef is het object in de lijst met als volgnummer het startgetal afgerond naar boven. Is het startgetal gelijk aan 5,00598, dan is het eerste object dus 6.

Voorbeeld 5.7. Bepaling van het startgetal

Voor een systematische steekproef van 500 adressen uit 9.590 adressen is de staplengte gelijk aan 19,18. U moet het *startgetal* loten uit het interval van 0 t/m 19,18.

Met de rekenmachine bepaalt u een willekeurige waarde tussen 0 en 1. Stel dat de uitkomst 0,261 is. Vermenigvuldig vervolgens die waarde met de staplengte 19,18. Dit levert dan het startgetal 5,00598.

Als de waarden van de staplengte en het startgetal bekend zijn, dan ligt daarmee de steekproef vast. Het eerste object in de steekproef is het object in de lijst met als volgnummer het startgetal afgerond naar boven. Is het startgetal gelijk aan 5,00598, dan is het eerste object dus 6.

Het volgende object krijgt u door bij het startgetal de staplengte op te tellen. Dat levert een nieuwe waarde. U rondt weer af naar boven en zo krijgt u het volgnummer van het tweede geselecteerde object. U gaat door met dit proces totdat u het einde van de lijst hebt bereikt. Dus het volgnummer van elk volgende object in de steekproef krijgt u steeds door bij de vorige waarde de staplengte op te tellen en af te ronden naar boven.

Voorbeeld 5.8. Trekken van een systematische steekproef

De lijst met adressen voor een luisteronderzoek van een lokale omroep bevat 9.590 adressen. Als u daaruit een systematische steekproef van 500 adressen wilt trekken, dan is de staplengte gelijk aan 19,18. Als het startgetal gelijk is aan 5,00598, dan krijgt u de volgende reeks waarden:

5,00598 24,18598 43,36598 . . . 9537,46598 9556,64598 9575,82598.

Als u al deze waarden afrondt naar boven, dan levert dat de volgende reeks volgnummers van de te selecteren objecten:

6 25 44 . . . 9538 9557 9576.

Als de deling van de lengte van de lijst door de omvang van de steekproef een geheel getal oplevert, dan zijn de berekeningen allemaal wat simpeler. Al dat afronden is dan niet nodig. Als u bijvoorbeeld een systematische steekproef van 500 objecten moet trekken uit een lijst van 9.500 objecten, dan is de staplengte gelijk aan $9500 / 500 = 19$. De startwaarde loot u dan uit de gehele getallen 1 t/m 19. Zou dit 5 opleveren, dan bestaat de steekproef uit de elementen 5, 14, 23, . . . , 9448, 9467, 9486.

Een waarschuwing is wel op zijn plaats bij het trekken van een systematische steekproef uit een lijst. Deze manier van trekken veronderstelt dat de volgorde van de objecten in de lijst volstrekt willekeurig is. In ieder geval moet de volgorde totaal niets te maken hebben met het onderwerp van het onderzoek.

Een simpel voorbeeld illustreert het gevaar dat bij systematisch trekking op de loer ligt. Stel een lijst bestaat uit namen en adressen van personen, waarbij alle personen uit een gezin achter elkaar staan. Bij bijvoorbeeld een staplengte van 20 zullen dan nooit meer personen uit één gezin in de steekproef kunnen komen. Dat betekent dat personen uit grote gezinnen altijd ondervertegenwoordigd zullen zijn, welke steekproef u ook trekt. Mochten grote gezinnen zich anders gedragen dan kleine gezinnen met betrekking tot het verschijnsel dat u onderzoekt, dan levert een dergelijke 'scheve' steekproef dus een verkeerd beeld op.

Als u een systematische steekproef trekt uit een lijst die gesorteerd is op straatnaam of postcode, of u trekt adressen uit een telefoonboek, dan mag u in de meeste gevallen veronderstellen dat dit soort vervelende effecten zich niet zullen voordoen.

5.5 De tweetrapssteekproef

Als u een steekproef van personen wilt trekken en u gebruikt daarvoor een steekproefkader van adressen, dan trekt u in feite een *tweetrapssteekproef*. In de eerste trap trekt u adressen, en in de tweede trap personen op de in de eerste trap geselecteerde adressen. De vraag is nu wie u moet enquêteren op de geselecteerde adressen: alle bewoners die tot de doelpopulatie behoren, of alleen maar de persoon die de deur open doet of de telefoon aanneemt, of misschien een willekeurige persoon op het adres?

Het is in veel situaties niet zo zinvol om meer mensen op hetzelfde adres te interviewen. Als meningen of gedragingen van de leden van een gezin op elkaar lijken, dan levert interviewen van meer dan één persoon op het adres niets nieuws op. Het is dan beter om de steekproef zoveel mogelijk verspreid over zoveel mogelijk verschillende adressen te trekken. Dit pleit er dus voor om maar één persoon per adres in de steekproef te trekken.

Het komt voor, maar niet zo vaak, dat op één adres verschillende huishoudens wonen. Daarbij kunt u bijvoorbeeld denken aan woningen voor studenten. In zo'n situatie is het gebruikelijk dat u eerst willekeurig één huishouden selecteert en vervolgens daaruit één willekeurig persoon. In feite is hier dan sprake van een *drietrapssteekproef*.

Hoe selecteert u nu één willekeurige persoon uit een huishouden? Daarvoor moet u eerst inventariseren hoeveel personen er op het adres wonen, en die bovendien tot de doelpopulatie behoren. Dat zou bijvoorbeeld kunnen betekenen dat u alleen leden van het huishouden boven een bepaalde leeftijd meetelt. Uit het groepje leden dat u zo krijgt, moet u één persoon selecteren. Bij een mondelinge of telefonische enquête moet de enquêteur dat doen. En bij een schriftelijke enquête of online-enquête moet één van de leden van het huishouden dat doen. Om problemen te voorkomen, moet de procedure simpel zijn. Een veel gebruikte procedure is de volgende: selecteer de persoon die het eerste jarig is. Daarbij is van de (niet onredelijke) veronderstelling uitgegaan dat er geen verband zal bestaan tussen de datum van de verjaardag en het onderwerp van de peiling.

Merk op dat door het selecteren van één persoon per huishouden de trekkingskansen van de personen in de doelpopulatie niet meer gelijk zijn. Ieder huishouden heeft wel een even grote kans om in de steekproef te komen (even afgezien van de mogelijkheid van meer huishoudens per adres), maar personen in grote huishoudens hebben een kleinere kans dan personen in kleine huishoudens. Bij het maken van schattingen op grond van de steekproef dient u voor deze ongelijke trekkingskansen te corrigeren. Doet u dat niet, dan loopt u het risico onjuiste conclusies te trekken. Personen uit grote huishoudens zijn immers ondervertegenwoordigd in de steekproef. Om te kunnen corrigeren is het belangrijk dat u het aantal leden (voor zover behorend tot de doelpopulatie) van elk geselecteerd huishouden registreert.

Voorbeeld 5.9. Trekkingskansen in een luisteronderzoek

In een luisteronderzoek wilt u op 209 adressen iemand interviewen. Het totaal aantal adressen in de doelpopulatie is gelijk aan 9.590. Daarmee is de trekkingskans van elk adres gelijk aan

$$\frac{209}{9590} = 0,022 .$$

Verder wordt de trekkingskans van een persoon op een geselecteerd adres bepaald door het aantal daar wonende personen van 12 jaar en ouder. Als we dit aantal aangeven met A , dan is die trekkingskans gelijk aan:

$$\frac{1}{A} .$$

De totale kans voor een persoon om in de steekproef te komen, krijgt u door beide bovenstaande kansen met elkaar te vermenigvuldigen. Dit geeft als trekkingskans:

$$\frac{209}{9590 \times A} .$$

Uit deze formule blijkt dat niet elke persoon in de doelpopulatie dezelfde kans heeft om in de steekproef terecht te komen. Een persoon in een 1-persoons huishouden heeft bijvoorbeeld een kans van $209 / 9590 = 0,022$, en de kans voor een persoon in een 2-persoons huishouden is $209 / 19180 = 0,011$ (tweemaal zo klein).

Hoe u op basis van de steekproef schattingen kunt maken van populatiekenmerken, is het onderwerp van het volgende hoofdstuk. Daarin leggen we ook uit hoe u de nauwkeurigheid van uw schattingen kunt bepalen. Verder proberen we in dat hoofdstuk ook de vraag te beantwoorden hoe groot de steekproef moet zijn.

Insluitkansen

De kans voor een object in de doelpopulatie om in de steekproef te worden getrokken, noemen we formeel de *insluitkans*.

Als u een enkelvoudige aselecte steekproef van omvang n trekt uit een doelpopulatie van omvang N , dan is de insluitkans gelijk aan n / N .

Als u eerst adressen trekt met gelijke kansen, en vervolgens op elke geselecteerd adres één persoon, dan is de insluitkans van een persoon gelijk aan $n / (N \times A_k)$, waarbij A_k het aantal leden van het huishouden is op adres k (voor zover ze tot de doelpopulatie behoren).

5.6 De praktijk

In deze paragraaf bespreken we een aantal mogelijkheden om in de praktijk een enkelvoudige aselecte steekproef zonder teruglegging te trekken. Uitgangspunt is dat we een steekproef willen trekken uit de bevolking van een gemeente. De mogelijkheden worden bepaald door wat er aan steekproefkaders beschikbaar is. Daarbij heeft elk in aanmerking komend steekproefkader zijn voor- en nadelen.

Steekproef uit het bevolkingsregister

Elk gemeente houdt een bevolkingsadministratie bij. Het systeem daarvoor heet de Gemeentelijke Basisadministratie Persoonsgegevens (GBA). In dit systeem staan alle inwoners van de gemeente met (onder andere) naam, adres en geboortedatum. Zeker bij de grotere gemeenten is het technisch mogelijk om een steekproef van personen uit deze administratie te (laten) trekken, en daarbij ook voorwaarden te stellen, bijvoorbeeld alleen mensen binnen zekere leeftijdsgrenzen.

Een steekproef uit het bevolkingsregister is alleen mogelijk als het gemeentebestuur daarvoor toestemming geeft. Bovendien kunnen de gemeenten kosten in rekening brengen voor deze vorm van dienstverlening.

De software voor het onderhoud van het GBA is niet bij elke gemeente hetzelfde. Het hoeft daarom niet zo te zijn dat de software faciliteiten biedt voor het trekken van een enkelvoudige aselecte steekproef. Ook is bij de gemeentelijke automatiseringsafdeling lang niet altijd de kennis, ervaring en capaciteit aanwezig om hiervoor zelf software te ontwikkelen. Een alternatief kan dan eventueel zijn het gebruik maken van het *A-nummer*. Dat is een uniek identificatienummer dat iedere persoon in het GBA heeft. Dit nummer wordt alleen intern gebruikt door de gemeenten. Het A-nummer is in wezen een aselect getal van 10 cijfers. Om een steekproef te trekken zou u aan de gemeente kunnen vragen om die personen te selecteren waarvan, bijvoorbeeld, het tweede cijfer van het A-nummer gelijk is aan 4. Zo krijg je een steekproef van 10% van de bevolking. Daaruit filtert u dan eerst de personen die aan de criteria voldoen. Is de steekproef dan nog te groot, dan kunt u hem verder uitdunnen of alleen het eerste deel ervan gebruiken.

Is de gemeente veel groter, dan kunt u ook besluiten eerst een steekproef van 1% te laten trekken uit het GBA door twee cijfers vast te leggen. Bijvoorbeeld: het derde cijfer moet gelijk zijn aan 8 en het vierde cijfer moet gelijk zijn aan 1.

Voorbeeld 5.10. Trekkingskansen in een luisteronderzoek

Stel, u moet een steekproef van 500 personen trekken uit een gemeentelijke populatie van 19.000 personen. Via het A-nummer trekt u een steekproef van 10%. Dat levert een lijst op met de namen en adressen van 1.900 personen. Vervolgens gebruikt u de eerste 500 personen in deze lijst.

Door het bevolkingsregister te gebruiken als steekproefkader, bestaat uw doel-populatie uit alle personen die staan ingeschreven bij de gemeente en aan een zekere leeftijdsgrenzen voldoen. Daarbij zitten ook mensen die tijdelijk in het buitenland verblijven. En daarbij zitten niet de mensen die tijdelijk in de gemeente verblijven, en ook niet de mensen die in de gemeente werken, maar ergens anders wonen.

Een steekproef uit een adressenbestand

Er zijn bedrijven die (tegen betaling) een *adressenbestand* kunnen leveren. Die bestanden zijn vaak gebaseerd op het postafgiftepuntenbestand van PostNL. U kunt eventueel ook alleen mensen selecteren die bepaalde kenmerken hebben. Selectie op leeftijd is bijvoorbeeld mogelijk. U moet dan nog wel per adres een of meer personen selecteren.

Merk op dat er sprake kan zijn van *overdekking*. Er kunnen bijvoorbeeld op een adres (tijdelijk) mensen wonen die niet staan ingeschreven in de gemeente. Het adressenbestand moet natuurlijk up-to-date zijn. Is dat niet het geval, dan kunnen er adressen instaan van huizen die ondertussen zijn afgebroken. En ook pas gebouwde huizen kunnen ontbreken.

Voorbeeld 5.11. Een fragment van een adressenbestand

Een voorbeeld van een kort stukje van een adressenbestand is hieronder weergegeven. Het betreft adresgegevens in de gemeente Leiderdorp.

2353	LK	23	Hella Haassestraat
2353	LK	24	Hella Haassestraat
2353	LK	25	Hella Haassestraat
2353	LL	1	Jan de Hartogplein
2353	LL	2	Jan de Hartogplein
2353	LL	3	Jan de Hartogplein
2353	LL	4	Jan de Hartogplein
2353	LL	4 B	Jan de Hartogplein
2353	LL	4 C	Jan de Hartogplein
2353	LL	4 D	Jan de Hartogplein
2353	LL	4 E	Jan de Hartogplein
2353	LL	4 F	Jan de Hartogplein
2353	LL	4 H	Jan de Hartogplein
2353	LL	4 J	Jan de Hartogplein

Elke regel bevat een adres. De regel begint met de postcode. Daarna volgen het huisnummer, met eventuele toevoegingen. En aan het einde van de regel staat de straatnaam. Van elk van de vermeldingen in het bestand is bekend of het een woonhuis betreft, een bedrijf, of iets anders. Daarom is het mogelijk te vragen een steekproef te trekken die, bijvoorbeeld, alleen uit adressen van woonhuizen bestaat.

Het is denkbaar om een adressenbestand op te bouwen uit een postcodebestand. In een postcodebestand is de eenheid niet het adres, maar de postcode. Bij elke postcode staat vermeld welke adressen erbij horen. Er kunnen wat haken en ogen zitten aan het ombouwen van een postcodebestand naar een adressenbestand. Zie voorbeeld 5.12.

Heeft u een postcodebestand omgezet in een adressenbestand, dan kunt u hieruit een steekproef trekken. Merk op dat er in ernstige mate sprake kan zijn van *overdekking*. U kunt adressen trekken waarop mensen wonen die niet in de gemeente staan ingeschreven. Er zullen ook veel adressen bijzitten die niet bij woonhuizen horen. Bij het vaststellen van de omvang van de te trekken steekproef dient u ermee rekening te houden dat een flink deel van de adressen onbruikbaar zal zijn. De steekproef waarmee u begint (de bruto steekproef) moet groter zijn dan de steekproef die u uiteindelijk wilt overhouden (de netto steekproef).

Voorbeeld 5.12. Een fragment van een postcodebestand

Een voorbeeld van een kort stukje van een postcodebestand is hieronder weergegeven. Het betreft postcodes in de gemeente Leiderdorp.

2353	LH	Maarten 't Hartstraat 1 t/m 23	Leiderdorp
2353	LH	Maarten 't Hartstraat 2 t/m 16	Leiderdorp
2353	LK	Hella Haassestraat 1 t/m 25	Leiderdorp
2353	LL	Hella Haassestraat 2 t/m 24	Leiderdorp
2353	LL	Jan de Hartogplein 1 t/m 57	Leiderdorp
2353	LL	Jan de Hartogplein 2 t/m 56	Leiderdorp

Elke regel bevat een postcode gevolgd door een reeks adressen die daarbij horen. Merk op dat de even en oneven huisnummers in aparte regels worden vermeld.

In principe lijkt het mogelijk om uit een dergelijk postcodebestand een adressenbestand voor de gehele gemeente te reconstrueren. Maar er kunnen problemen optreden. Zo is bijvoorbeeld niet te zien dat niet alleen het adres Jan de Hartogplein 4 voorkomt, maar ook nog eens de adressen Jan de Hartogplein 4b t/m Jan de Hartogplein 4j. Om te voorkomen dat adressen met huisnummers met toevoegingen uitgesloten worden bij de steekproeftrekking, zult u dit probleem op een of andere manier moeten oplossen.

Een steekproef uit het telefoonboek

Een derde mogelijkheid om een steekproef te trekken is gebruik maken van het telefoonboek van de gemeente. Dat zou handig kunnen zijn als u een telefonische enquête wilt uitvoeren. Het telefoonboek bevat echter ook adresgegevens, zodat u dit steekproefkader ook zou kunnen overwegen voor een mondelinge of schriftelijke enquête.

Het telefoonboek is beslist geen ideaal steekproefkader. Er is een aantal tekortkomingen:

- Niet elk gezin heeft een vaste telefoon (onderdekking). Vooral steeds meer jongeren hebben een mobiele telefoon. De mobiele nummers staan niet in het telefoonboek.
- Niet elk gezin met een vaste telefoon staat vermeld in het telefoonboek (onderdekking).
- De vermelde nummers horen vaak niet bij een gezin. Een groot deel van de nummers is van bedrijven en instellingen (overdekking).
- Veel mensen hebben zich tegenwoordig geregistreerd in het Bel-me-niet Register. Dat betekent dat u ze niet mag lastig vallen met telefonische verkooppraatjes. U kunt zich afvragen of het dan verstandig is om ze te bellen voor een enquête. Dat mag omdat het niet om verkoop gaat, maar zal men het verschil zien?

Het trekken van een steekproef uit het telefoonboek gaat het eenvoudigst met een systematische trekking. Daarvoor moet u eerst tellen hoeveel nummers er in totaal voor de gemeente in het telefoonboek staan. Vervolgens kunt u de trekkingsprocedure uit paragraaf 4.4 toepassen.

U kunt de systematische trekkingsprocedure eventueel nog iets verder vereenvoudigen door de trekking op het niveau van pagina's of kolommen uit te voeren.

Omdat er veel overdekking zit in het telefoonboek, is het belangrijk meer nummers te trekken dan nodig zijn. Voor sommige nummers zal het uit de vermelding al meteen duidelijk zijn dat het niet om een woonhuis gaat. Maar bij andere nummers zal dat pas duidelijk worden als u via de telefoon contact legt. Alvorens het interview te starten moet u dus eerst vaststellen of het om woning of om een bedrijf gaat.

Voorbeeld 5.12. Steekproef uit het telefoonboek

Voor de gemeente Leiderdorp stonden ooit ongeveer 11.000 nummers vermeld in het telefoonboek. Die waren verdeeld over 46 pagina's met elk 4 kolommen. Er waren dus 184 kolommen. Voor een steekproef van 1.000 nummers kunt u dus uit elke kolom 6 nummers trekken.

6. Het verzamelen van gegevens

Voor elk object in de steekproef moet een vragenlijst worden ingevuld. Is het object, de onderzoekseenheid, een persoon, dan ligt het voor de hand wie de vragenlijst moet invullen. Is de onderzoekseenheid een huishouden, dan zal iemand in het huishouden het moeten doen. En voor een peiling bij bedrijven zal voor elk geselecteerd bedrijf een vertegenwoordiger ervan het formulier moeten invullen. Bij ingewikkelde bedrijfsonderzoeken kan het zelfs gebeuren dat meer medewerkers elk een deel van de vragenlijst invullen. In dit hoofdstuk ligt de nadruk vooral op peilingen bij personen.

Hoe gaat u de vragenlijsten laten invullen? Dat kan op verschillende manieren:

- *Mondeling*: enquêteurs gaan bij de geselecteerde personen op bezoek. De enquêteur stelt de vragen en de personen geven de antwoorden,
- *Telefonisch*: enquêteurs bellen de geselecteerde personen. De enquêteur stelt telefonisch de vragen en de personen geven de antwoorden,
- *Schriftelijk*: U stuurt de vragenlijst in een brief naar de geselecteerde personen. Die personen vullen het formulier zelf in en sturen het daarna terug.
- *Online*: U stuurt de selecteerde personen een link naar een website met de elektronische vragenlijst. Die personen vullen de vragenlijst zelf in op hun computer.

Bij de keuze van de methode van gegevensverzameling spelen allerlei aspecten een rol. De belangrijkste twee aspecten zijn misschien wel kosten en kwaliteit. Als u enquêteurs inzet, dan levert dat meestal goede gegevens op, maar de kosten zijn hoog. Als u het zonder enquêteurs doet, is het goedkoper, maar u betaalt daar een prijs voor in termen van de kwaliteit van de antwoorden.

Het verzamelen van de gegevens noemen we ook nog wel het *veldwerk*. Die term verwijst eigenlijk naar peilingen waarbij enquêteurs op pad gaan ('het veld in gaan') om bij de respondenten thuis de vragenlijsten in te vullen. We gebruiken de term hier ook voor andere manieren van het verzamelen van de gegevens.

In dit hoofdstuk beschrijven we verschillende methoden van gegevensverzameling. We beginnen met drie traditionele manieren van het verzamelen van gegevens. Daarbij maakt u gebruik van papieren vragenlijsten. Dan beschrijven we enkele manieren van computergestuurd enquêteren. Hierbij is de papieren vragenlijst vervangen door een elektronische vragenlijst in een computerprogramma. Ten slotte kijken we nog naar enquêteren via het internet.

6.1 Traditionele gegevensverzameling

Bij *schriftelijke enquêteren* verstuurt u de vragenlijst per post naar de personen in de steekproef. Daarbij zit het uiteraard het verzoek de vragenlijst ingevuld terug te

sturen. U hoeft geen enquêteurs in te zetten. Daarom is dit een goedkope manier van gegevensverzameling. Een bijkomend voordeel is het ontbreken van de bedreigende werking die soms van het bezoek van een (onbekende) enquêteur uitgaat. Verder worden gevoelige vragen beter beantwoord als er geen enquêteur bij is. De afwezigheid van enquêteurs heeft echter ook nadelen. Er is niemand die de persoon kan overtuigen van het belang van het invullen van de vragenlijst. En ook kan de enquêteur niet helpen bij het invullen van de vragenlijst. Verder moet u hoge eisen stellen aan de vraagstelling in en opmaak van de vragenlijst. De respondenten moeten hem zonder problemen kunnen invullen. Doordat schriftelijk enquêteren de indruk wekt nogal vrijblijvend van karakter te zijn, kan dit resulteren in een hoge non-respons. Veel mensen zullen de vragenlijst niet invullen, maar bij het oud papier gooien.

Vanwege de nadelen van schriftelijk enquêteren zou u de voorkeur kunnen geven aan *mondeling enquêteren*. Hierbij bezoeken enquêteurs de geselecteerde personen thuis en proberen ze in een persoonlijk gesprek de vragen beantwoord te krijgen. Een punt van overweging is of u de peiling van te voren moet aankondigen in de media. U kunt dan uitleggen wat het doel van je onderzoek is, dat het belangrijk is om mee te doen, en dat u de gegevens vertrouwelijk zult behandelen. Het is bij veel onderzoek ook niet ongebruikelijk om het bezoek van een enquêteur aan te kondigen in een *aanschrijfbrief*. Het moet er allemaal toe leiden dat mensen mee doen aan het onderzoek. Een nadeel van mondeling enquêteren is dat het duur is. U moet een team enquêteurs opleiden en betalen. Bovendien zijn er nog aanzienlijke reiskosten. Mondeling enquêteren levert echter over het algemeen een hogere respons op, terwijl ook de verzamelde gegevens van een goede kwaliteit zijn.

Een methode die een compromis vormt tussen schriftelijk en mondeling enquêteren is het *telefonisch enquêteren*. U hebt dan ook wel enquêteurs nodig, maar minder dan bij mondeling enquêteren. Die enquêteurs hoeven niet te reizen, wat een aanzienlijke tijds- en kostenbesparing oplevert. Daar staat tegenover dat de gestelde vragen niet te ingewikkeld mogen zijn. En als het gesprek te lang duurt, dan haken de respondenten af.

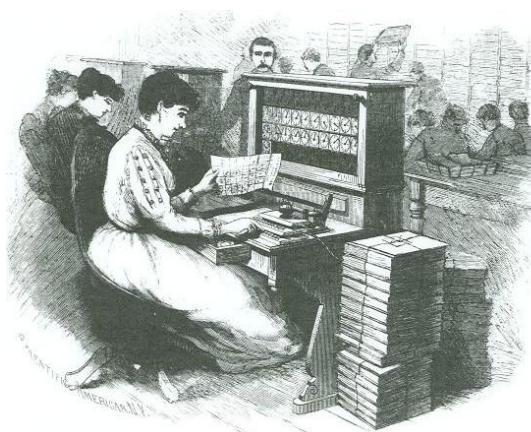
Als u telefonisch enquêteren overweegt, moet u zich wel afvragen of er een geschikt steekproefkader is. In het vorige hoofdstuk is al uitgelegd, dat heel veel mensen niet in het telefoonboek staan.

De keuze van de manier van gegevensverzameling blijft een lastige zaak. Vaak zal de keuze een compromis zijn, waarbij u aan de ene kant probeert de kosten zo laag mogelijk te houden, en andere kant de kwaliteit zo hoog mogelijk.

6.2 Computergestuurd enquêteren

De computer heeft altijd een grote rol gespeeld bij de verwerking van statistische gegevens. Al in 1890 gebruikte het Amerikaanse Census Bureau zogenaamde Hollerith-machines voor het verwerken van de formulieren van de volkstelling.

Figuur 6.1. Een Hollerith-machine



Pas in de laatste decennia heeft de computer ook zijn intrede gedaan bij het enquêteren zelf. Dat werd mogelijk door de opkomst van de relatief goedkope en gebruikersvriendelijke microcomputers. Essentieel element van computergestuurd enquêteren is dat de vragenlijst niet op papier staat, maar in een computerprogramma zit. Dit programma bepaalt welke vragen moeten worden gesteld en het controleert ook de antwoorden.

Het gebruik van computers tijdens het verzamelen van de gegevens heeft drie belangrijke voordelen. In de eerste plaats ontlast u de enquêteurs tijdens het interview. Die hoeven niet steeds de volgende vraag op te zoeken. Dat doet de computer voor hen. Dat lijkt misschien een simpele handeling, maar soms hangt het stellen van een vraag af van de antwoorden op al eerder gestelde vragen. U moet dan naar een ander deel van de vragenlijst springen. Een fout is hierbij snel gemaakt.

In de tweede plaats kan de computer tijdens het gesprek allerlei controles uitvoeren. Fouten die u tijdens het gesprek vindt, kunt u ook tijdens het gesprek verbeteren. Dat gaat beter tijdens het interview dan achteraf. Als u de respondent tijdens het gesprek kunt confronteren met een fout, dan kan hij snel worden verbeterd. Als u achteraf merkt dat er een fout in een formulier zit, dan is het vrijwel niet te doen om opnieuw contact op te nemen met de respondent in een poging de fout te verbeteren.

Het derde voordeel van het gebruik van de computer is dat na afloop van het gesprek de gegevens meteen al in de computer zitten. Dat hoeft u dus niet meer achteraf te doen. Dat levert een aanzienlijke versnelling op bij het verwerken van de gegevens.

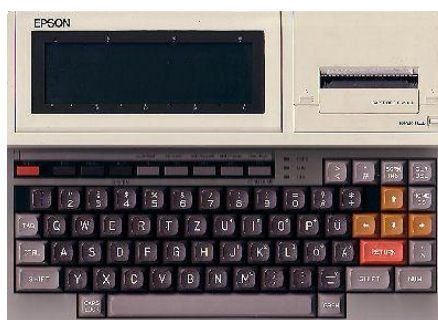
De computer werd het eerst ingezet bij telefonisch enquêteren. Dat gebeurde in de jaren 70 van de vorige eeuw. De enquêteurs hebben de beschikking over een telefoon en een computer. Ze nemen telefonisch contact op met de respondenten en starten vervolgens het interview. De computer stuurt het gesprek en controleert de antwoorden. Deze vorm van enquêteren wordt algemeen aangeduid met de term *CATI (Computer Assisted Telephone Interviewing)*.

Meer recent is het gebruik van de computer bij mondeling enquêteren. Net zoals dat het geval is bij CATI, zit ook hier weer de vragenlijst in de computer. Alleen betreft het hier een laptop of tablet. Dit zijn handzame computers die door de enquêteurs makkelijk zijn mee te nemen naar de te bezoeken personen. Daar neemt de computer

weer het heft in handen bij het stellen van de vragen en het controleren van de antwoorden. Mondeling enquêteren duiden we meestal aan met de term *CAPI* (*Computer Assisted Personal Interviewing*).

Het CBS begon al in 1984 te experimenteren met shootcomputers. Het bleek dat de enquêteurs goed met het nieuwe medium overweg konden. Ook bleek dat de respondenten geen bezwaren maakten tegen het gebruik van computers voor enquêteren. Er waren geen aanwijsbare psychologische ('Big Brother') effecten.

Figuur 6.2. De Epson PX-4 notebook-computer



In 1987 begon het CBS met CAPI in de Enquête Beroepsbevolking (EBB). Zo'n 400 enquêteurs kregen een laptop. Dat was toen een EPSON PX-4, draaiend onder het besturingssysteem CP/M. Iedere maand bezochten enquêteurs ongeveer 12.000 adressen en namen daar in totaal pakweg 30.000 interviews af. Veel meer over de introductie van computers voor enquêteren bij het CBS kunt u lezen in CBS (1987).

Ook schriftelijk enquêteren wordt steeds meer geautomatiseerd. De gebruikte terminologie doortrekkend, kunt u deze vorm van gegevensverzameling aanduiden met *CASI* (*Computer Assisted Self Interviewing*). Bij CASI wordt de vragenlijst in elektronische vorm naar de respondenten toegestuurd. De respondent start op zijn eigen computer het enquêteerprogramma en typt de antwoorden in op de door het programma gestelde vragen. Daarna worden de aldus verzamelde gegevens ook weer op elektronische wijze teruggestuurd. Deze wijze van dataverzameling wordt vooral toegepast bij bedrijfsenquêtes.

Je kunt je afvragen of CASI een goede term is voor deze manier van gegevensverzameling. Er komt immers geen interviewer aan te pas en dus is er geen sprake van 'Interviewing'. We interpreteren het hier echter als respondenten die zichzelf interviewen.

Een van de eerste toepassingen van CASI bij personen thuis was het *Telepanel*. Dit werd in 1986 bij de Universiteit van Amsterdam ontwikkeld door Willem Saris. Hij zette een panel van huishoudens op. De leden van die huishoudens moesten regelmatig (eens per week) een vragenlijst invullen. Daarvoor kregen de geselecteerde huishoudens thuis een simpele homecomputer. Ze verbonden die met hun TV. Eens per week ontvingen ze per telefoon en modem een nieuwe elektronische vragenlijst. Ze moesten die op de computer invullen. Na voltooiing van de vragenlijst stuurden ze de antwoorden ook weer per telefoon en modem op terug. Voor meer informatie over het Telepanel, zie Saris (1998).

Computergestuurd enquêteren levert dus betere gegevens op. Er zijn echter ook wel nadelen. Zo kost de aanschaf van computers natuurlijk veel geld. En ook moet u software kopen of zelf ontwikkelen waarmee u de vragen elektronisch kunt stellen. Couper et al. (1998) geven een aardig overzicht van de rol van de computer bij peilingen.

6.3 Peilingen via het internet

Met de razendsnelle opkomst van het internet heeft een nieuw type gegevensverzameling zijn intrede gedaan: peilingen via het internet. Dit wordt soms ook wel *CAWI (Computer Assisted Web Interviewing)* genoemd. Hierbij biedt u de vragenlijst aan via het internet en de respondent interviewt zichzelf. Je zou een internet-enquête kunnen zien als een speciale vorm van CASI.

6.3.1 De populariteit van peilen via het internet

Op het eerste gezicht heeft enquêteren via het internet een aantal aantrekkelijke eigenschappen:

- U krijgt op een betrekkelijk eenvoudige manier toegang tot een zeer grote groep potentiële respondenten, namelijk iedereen met internet.
- U kunt de vragenlijsten tegen zeer lage kosten aanbieden. U hoeft immers geen enquêteurs in te schakelen, u hebt geen drukkosten, en u hebt ook geen verzendkosten.
- U kunt een peiling heel snel uitvoeren. Er hoeft maar weinig tijd verloren te gaan tussen opstellen en aanbieden van de vragenlijst.
- Het internet biedt aantrekkelijke extra mogelijkheden om zaken als beeld (foto, video, animatie) en geluid in de vragenlijst op te nemen.

6.3.2 Hoe representatief zijn peilingen via het internet?

Een peiling via het internet lijkt een snelle, goedkope en aantrekkelijke manier om veel gegevens te verzamelen. Het is echter niet allemaal rozengeur en maneschijn. De relatieve eenvoud waarmee we een internetpeiling kunnen opzetten, leidt tot een groot en nog steeds groeiend aanbod van dit soort peilingen. Er zijn websites (zoals *SurveyMonkey* and *LimeSurvey*) waarmee je in korte tijd een enquête in de lucht kunt brengen, ook al heb je geen enkel verstand van survey-methodologie. Veel van die enquêtes zijn niet op verantwoorde wijze opgezet. En door het grote aanbod is het moeilijk het kaf van het koren te scheiden.

De veelheid aan internetpeilingen draagt ook het gevaar in zich van een steeds meer toenemende non-respons. Potentiële respondenten worden overvoerd en haken af. In feite treedt hetzelfde effect op als dat waaraan telefonische peilingen lijden: de grote hoeveelheid aan telefonische verkoopactiviteiten waarmee mensen worden lastig gevallen, vermindert de bereidheid om aan echte telefonische enquêtes mee te werken.

Samenvattend kunnen we zeggen dat veel internetpeilingen van het type zijn waarbij u tegen lage kosten en in korte tijd een grote hoeveelheid gegevens verzameld. De grote nadruk op deze aspecten staat in veel gevallen op gespannen voet met zaken als de validiteit van de uitkomsten.

Bij het overwegen van een internetpeiling moeten u goed letten op een aantal zaken. Een van de belangrijkste problemen van dit soort peilingen is misschien wel de *onderdekking*. Hoewel steeds meer mensen toegang hebben tot het internet, is het zeker niet zo dat iedereen dat heeft. Dat betekent dat vooral ouderen, laagopgeleiden en allochtonen vaak onvoldoende vertegenwoordigd zullen zijn. De peiling is dus niet representatief.

U zou het probleem van de onderdekking kunnen oplossen door mensen zonder internet de mogelijkheid te geven om de vragenlijst op een andere manier in te vullen. U zou deze mensen bijvoorbeeld een papieren vragenlijst kunnen sturen. Omdat we dan tegelijkertijd verschillende methoden van gegevensverzameling gebruiken, noemen we een dergelijke peiling ook wel een *mixed-mode peiling*.

Een ander probleem wordt gevormd door het trekken van de steekproef. Hoe trekt u een steekproef van internetgebruikers? Er is geen lijst met alle e-mailadressen beschikbaar. Veel internetpeilers omzeilen dit probleem door de internetgebruikers zichzelf te laten selecteren voor de peiling. We noemen dit *zelfselectie*. Via 'banners', 'popup windows' en andere vormen van reclame maken ze internetgebruikers attent op de peiling. De peiler hoopt dan dat deze gebruikers zich hierdoor laten overhalen.

Typische voorbeelden van zelfselectie zijn te vinden bij sommige politieke peilingen in Nederland. In deze panels zitten vooral mensen die het leuk vinden om aan peilingen mee te doen en die geïnteresseerd zijn in politiek. Daardoor kun je grote vraagtekens zetten bij de representativiteit van dit soort peilingen.

Het probleem van zelfselectie is dat je als onderzoeker de trekkingskansen niet meer in de hand hebt. Je kunt zelfs achteraf niet meer bepalen wat die trekkingskansen zijn geweest. Het is daarom onmogelijk om op basis van aldus verkregen gegevens betrouwbare schattingen te maken van populatiekenmerken.

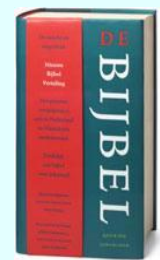
U kunt het probleem van de zelfselectie oplossen door de selectie van respondenten niet via het web te doen. U kunt, bijvoorbeeld, eerste een steekproef van namen en adressen van personen trekken. Die stuurt u vervolgens een brief waarin u ze uitnodigt naar de website te gaan waar de vragenlijst is te vinden. In die brief moet dan ook een unieke code staan waarmee de geselecteerde personen toegang krijgen tot de vragenlijst. Daarmee kunt u voorkomen dat iemand meer dan één keer de vragenlijst invult, en ook dat iemand van de buiten de doelpopulatie de vragenlijst invult.

Voorbeeld 6.1. De NS Publieksprijs

De NS publieksprijs is een literaire prijs die elk jaar wordt toegekend. Voor deze prijs wordt geen speciale jury ingesteld. De lezers kunnen zelf stemmen.

In 2005 werden zes boeken genomineerd: De thuiskomst van Anna Enquist, Troost van Ronald Giphart, Knielen op een bed violen van Jan Siebelink, De reünie van Simone van der Vlugt, Je gaat het pas zien als je het doorhebt van Pieter Winsemius en Sonny Boy van Annejet van der Zijl.

In een internetpeiling konden de lezers stemmen op een van deze zes boeken, maar ook zelf een boek opgeven als hun favoriete boek niet in de lijst stond. In totaal brachten 92.000 mensen hun stem uit. Tot verbazing van iedereen werd niet een van de genomineerde boeken tot winnaar gekozen. Zo'n 72% van stemmers kozen voor de Nieuwe Bijbelvertaling.



Deze verpletterende uitslag was het resultaat van een campagne gevoerd door onder anderen het dagblad Trouw, de Evangelische Omroep, het Nederlands Bijbelgenootschap, de Katholieke Bijbelstichting en de Protestantse Kerk om te stemmen op de nieuwe Bijbelvertaling.

Deze gang van zaken was niet tegen de regels van deze literaire prijs, maar je kunt je natuurlijk wel afvragen of deze uitslag representatief is voor de Nederlandse bevolking.

Omdat u bij een internetpeiling geen enquêteurs inzet, kan er van alles misgaan bij de beantwoording van de vragen. Gebrek aan motivatie bij de respondenten en verkeerd begrijpen van vragen kan aanleiding geven tot foutieve antwoorden. Die fouten kunnen ook worden veroorzaakt door slecht geformuleerde vraagteksten of een slecht ontwerp van de vragenlijst of technische problemen bij de beantwoording van de vragen.

Problemen met het ontwerp van een vragenlijst kunnen nog worden vergroot doordat de weergave ervan van respondent tot respondent kan verschillen, als gevolg van gebruik van verschillende browsers, van verschillende instellingen van dezelfde browser, of van verschillende apparaten (computer, netbook, tablet, mobiele telefoon). Kortom, grote voorzichtigheid is geboden bij het uitvoeren van een internetpeiling.

Respondenten zijn vaak niet echt geïnteresseerd in het onderwerp van het peiling. Daarom zullen ze niet hun uiterste best doen om de vragen correct en volledig te beantwoorden. Bij het ontwerpen van de vragenlijst is het daarom verstandig er vanuit te gaan dat

- de respondenten weinig belangstelling hebben voor het onderzoek,
- meedoen daarom niet belangrijk voor hen is,
- ze de vragen niet nauwkeurig lezen maar slechts globaal scannen,

- niet het beste antwoord kiezen, maar het eerste redelijk acceptabele antwoord dat ze tegenkomen,
- ze weten dat er geen straf staat op het geven van onjuiste antwoorden,
- ze niet uitzoeken hoe de vragenlijst werkt, maar al voortmodderend proberen het eind te bereiken.

6.3.3 Vragenlijsten op het internet

De opmaak van de verschillende typen vragen op het internet wordt voor een groot deel bepaald door wat mogelijk is in HTML, de taal om webpagina's te maken. Soms zijn er verschillende mogelijkheden, elk met zijn voor- en nadelen.

Figuur 6.3. Een gesloten vraag met één antwoord

Naar welke soort muziek heeft u de afgelopen zeven dagen het meest geluisterd?

- Pop
- Dance
- Rock
- R & B
- Hip-hop
- Country
- Folk
- Easy listening
- Jazz
- Klassiek
- New age
- Andere muziek

De meest voor de hand liggende manier om een gesloten vraag (met één antwoord) te maken is met *keuzerondjes* (*radio buttons*). Zie figuur 6.3 voor een voorbeeld. Daarbij kunt u inderdaad maar hooguit één antwoord aanklikken. Het aanklikken van een antwoord maakt de eerdere selectie van een ander antwoord weer ongedaan.

Een andere manier om een gesloten vraag te maken is met een *keuzelijst* (*drop down list*). Dat is een lijst die u eerst moet openklappen. Vervolgens kunt u door de lijst bladeren en op het juiste antwoord klikken. In vergelijking met de keuzerondjes zijn er bij een keuzelijst meer handelingen nodig om een antwoord te kiezen: lijst openen, door de lijst bladeren en antwoord aanklikken. De linker lijst in figuur 6.4 is nog gesloten. De middelste lijst toont de toestand na openen en bladeren.

Figuur 6.4. Een gesloten vraag met een keuzelijst

Naar welke soort muziek heeft u de afgelopen zeven dagen het meest geluisterd?

Klik hier voor het antwoord ▾

Naar welke soort muziek heeft u de afgelopen zeven dagen het meest geluisterd?

Klik hier voor het antwoord ▾

- Pop
- Dance
- Rock
- R & B
- Hip-hop
- Country
- Folk
- Easy listening
- Jazz
- Klassiek
- New age
- Andere muziek

Naar welke soort muziek heeft u de afgelopen zeven dagen het meest geluisterd?

- Pop
- Dance
- Rock
- R & B
- Hip-hop

Als de lijst van mogelijkheden lang is, dan zijn die niet allemaal tegelijk zichtbaar. In principe hangt het van de browser af hoeveel antwoorden tegelijk zichtbaar zijn. Het is echter mogelijk om het maximaal aantal zichtbare antwoorden in te stellen. In de rechterlijst in figuur 6.4 is dit aantal op 5 gezet.

Het beperkte aantal zichtbare antwoorden is een ernstig nadeel van de keuzelijst. De eerste keer dat u bij een dergelijke vraag komt, ziet u alleen de eerste paar mogelijke antwoorden. Dit leidt tot een *primacy effect* (voorkeur voor antwoord vooraan in de lijst).

Er zijn ook gesloten vragen waarbij het is toegestaan of meer dan één antwoord te selecteren. Het ligt voor de hand om dit soort vragen in HTML te maken met *aankruisvakjes* (*check boxes*). Hierbij selecteert u een antwoord door op het desbetreffende (vierkante) vakje te klikken. Er verschijnt dan een vinkje. Een antwoord blijft geselecteerd totdat u er weer op klikt.

De linker vraag in figuur 6.5 is gemaakt met aankruisvakjes. In het voorbeeld zijn drie antwoorden geselecteerd. Vooral als de lijst van mogelijke antwoorden erg lang is, kan dit leiden tot primacy effecten. En ook komt *satisficing* voor: de respondent kiest niet alle voor hem relevante antwoorden, maar selecteert een paar makkelijke antwoorden in de lijst en vindt het dan wel genoeg.

Figuur 6.5. Een gesloten vraag met meer mogelijke antwoorden

Naar welke soorten muziek heeft u de afgelopen zeven dagen het meest geluisterd? Er zijn meer antwoorden mogelijk.

- Pop
- Dance
- Rock
- R & B
- Hip-hop
- Country
- Folk
- Easy listening
- Jazz
- Klassiek
- New age
- Andere muziek

Naar welke soorten muziek heeft u de afgelopen zeven dagen het meest geluisterd? Er zijn meer antwoorden mogelijk.

Ja Nee

- Pop
- Dance
- Rock
- R & B
- Hip-hop
- Country
- Folk
- Easy listening
- Jazz
- Klassiek
- New age
- Andere muziek

Een mogelijke manier om satisficing te verminderen is het vormgeven van de vraag zoals in het voorbeeld rechts in figuur 6.5. Voor ieder mogelijk antwoord zijn er twee keuzerondjes: voor *ja* en voor *nee*. Je moet altijd een van die twee aanklikken. Je moet dus iets doen voor elk mogelijk antwoord. Dat is meer werk voor de respondent, maar experimenten lijken aan te tonen dat ze zo wel meer antwoorden selecteren en niet te snel stoppen.

Bij open vragen kan de respondent elke willekeurige tekst als antwoord invullen. Open vragen hebben niet de voorkeur omdat interpretatie en analyse van dit soort antwoorden problemen kunnen opleveren. Kies daarom alleen voor een open vraag als je de vraag niet in de vorm van een gesloten vraag kunt gieten.

Figuur 6.6 toont twee manieren om een open vraag vorm te geven in een online-vragenlijst. De eerste maakt gebruik van een *tekstveld*. U kunt dan één regel tekst

invoeren. De tweede maakt gebruik van een *tekstvak*. Hier kunt u meer regels tekst kwijt. De scrollbar aan de rechterkant geeft aan dat u nog veel meer tekst kunt invoeren dan zichtbaar is.

Figuur 6.6. Open vragen

1. Waarom heeft u gekozen voor deze universiteit?

2. Uw antwoord op deze vraag is voor ons erg belangrijk. We krijgen daardoor meer inzicht in waarom mensen naar deze universiteit komen. Waarom hebt u gekozen voor deze universiteit?

Uit onderzoek is gebleken dat het gebruik van een tekstveld leidt tot veel kortere antwoorden dan een tekstvak. Het is daarom van belang dat de vormgeving van de vraag overeenkomt met de hoeveelheid tekst die u als antwoord verwacht. Wilt u korte, maar krachtige antwoorden, gebruik dan een tekstveld. Maar wilt u veel tekst, gebruik dan een tekstvak.

Als een reeks gesloten vragen dezelfde antwoordmogelijkheden heeft, dan zou u kunnen overwegen ze te combineren in een *matrixvraag*. Elke rij in de matrix correspondeert dan met een vraag en elke kolom met een mogelijk antwoord. Figuur 6.7 bevat een voorbeeld.

Matrixvragen kom je vaak tegen in internetpeilingen. De achterliggende gedachte is dat een dergelijke vraag minder ruimte inneemt dan een reeks losse vragen. Daarom zou de vragenlijst overzichtelijker zijn dan een stel losse vragen.

Figuur 6.7. Een matrixvraag

	Uitstekend	Heel goed	Goed	Redelijk	Slecht
1. Wat vindt u in het algemeen van de kwaliteit van de omroep?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Wat vindt u van de kwaliteit van de nieuwsprogramma's?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Wat vindt u van de kwaliteit van de sportprogramma's?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Wat vindt u van de kwaliteit van de muziekprogramma's?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Wat vindt u van de kwaliteit van de culturele programma's?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Wat vindt u van de kwaliteit van de kinderprogramma's?	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Op het eerste gezicht lijken matrixvragen wel voordelen te hebben. Een matrixvraag neemt minder ruimte in dan een reeks losse vragen. Een dergelijke vraag biedt de respondenten ook wat meer overzicht. Daarom zou het beantwoorden van de vragen wel eens minder tijd kunnen kosten. Zoals eerder gemeld, zitten er haken en ogen aan het gebruik van een matrixvraag. Het is cognitief lastiger dan het beantwoorden van een simpele vraag. En ook bestaat het gevaar van *straight-lining*. Bij peilingen via het internet speelt ook nog mee dat de matrix van vragen maar voor een deel zichtbaar is op het scherm als gevolg van een lagere schermresolutie of schermomvang. Daardoor kun je antwoordmogelijkheden of vragen over het hoofd zien.

Veel deskundigen adviseren het gebruik van matrixvragen zoveel mogelijk te beperken. Als het dan toch moet, doe het dan zo goed mogelijk. Wat de respondent bijvoorbeeld helpt, is het geven van een afwisselend achtergrondkleur aan de rijen, zoals in figuur 6.7.

U stelt vragen in een enquête om informatie te verzamelen over de respondenten. Dan moeten de respondenten die vragen wel kunnen beantwoorden. Het komt echter voor dat een respondent het antwoord op een vraag niet weet. Om hiermee rekening te houden, zou u een speciale antwoordcategorie “Weet niet” kunnen opnemen. Het gevaar bestaat dan echter dat mensen die geen zin hebben om te antwoorden, uit gemakzucht ook voor “Weet niet” kiezen. Dit is een vorm van satisficing. De vraag is nu hoe om te gaan met ‘Weet niet’ in online-vragenlijsten. Er is een aantal verschillende vormen denkbaar:

- Biedt “Weet niet” expliciet aan als een van de mogelijke antwoorden. Hiermee accepteert u dat sommige respondenten het antwoord niet weten. Er treedt hierbij wel satisficing op: mensen antwoorden “Weet niet” terwijl ze het toch wel weten.
- Biedt “Weet niet” expliciet aan, maar op een minder duidelijke manier, bijvoorbeeld op een andere plek op het scherm of met kleinere of minder heldere tekst. Dit blijkt niet te werken. Het percentage “Weet niet” blijft te hoog. Er zijn ook respondenten die deze optie over het hoofd zien en klagen dat de optie “Weet niet” niet aanwezig is.
- Biedt “Weet niet” impliciet aan. De eerste keer dat de vraag op het scherm verschijnt, is er geen optie “Weet niet”. Maar als de respondent probeert de vraag over te slaan, verschijnt de vraag opnieuw, maar nu met de optie “Weet niet”. Dit leidt tot een lager percentage “Weet niet”, maar ook tot klachten dat in eerste instantie “Weet niet” niet te vinden is. Een ander vorm is die waarbij de vraag niet de optie “Weet niet” heeft, en waarbij een poging tot overslaan van de vraag leidt tot een keuze om de vraag alsnog te beantwoorden of “Weet niet” als antwoord te geven. Ook dit leidt tot een lager percentage “Weet niet”.
- Biedt “Weet niet” niet aan als mogelijk antwoord. Respondenten moeten dan een ‘echt’ antwoord geven. Dit vermindert satisficing, maar maakt het moeilijk voor respondenten die het echt niet weten. Diverse deskundigen zijn tegen deze vorm van “Weet niet”. Ze vinden dat je respondenten niet mag forceren tot een ‘verkeerd’ antwoord. Dat leidt tot frustratie bij respondenten en mogelijk tot afbreken van het invullen van de vragenlijst.

Bij computergestuurd enquêteren (CAPI en CATI) is het niet ongebruikelijk dat er allerlei controles in de vragenlijst zitten. Wanneer een respondent inconsistente antwoorden geeft, dan meldt de software dit en kunnen de gerapporteerde fouten worden verbeterd. Deze controle/correctie leidt tot gegevens van betere kwaliteit. De vraag is nu of u zulke controles ook in een online-vragenlijst moet opnemen. Dit zou betekenen dat de respondent (en niet de enquêteur) een foutmelding op het scherm

krijgt. De vraag is hoe deze daarop reageert. De terechtwijzing kan irritatie oproepen. En dat zou weer kunnen leiden tot non-respons. U wordt geconfronteerd met een dilemma om te kiezen voor betere gegevens met een lagere respons of slechtere gegevens en een hogere respons. Mocht u besluiten tot het inbouwen van controles van de antwoorden, dan moeten de foutmeldingen en waarschuwingen in ieder geval zo vriendelijk mogelijk zijn.

Meer over de mogelijkheden van het peilen via het internet, maar ook over de voetangels en klemmen, kunt u lezen in Bethlehem & Biffignandi (2012). Andere nuttige boeken over peilen via het internet zijn Couper (2008) en Dillman, Smyth & Christian (2009).

7. Controle en correctie

Na afloop van het veldwerk heeft u een grote hoeveelheid ingevulde formulieren. Alle informatie op die formulieren moet u analyseren. Daarmee kunt u echter niet meteen beginnen. Helaas is het zo dat enquêteurs en respondenten fouten maken bij het beantwoorden van de vragen en het invullen van de formulieren. Om te voorkomen dat u incorrecte gegevens analyseert en daardoor misschien wel verkeerde conclusies trekt uit uw onderzoek, dient u de verzamelde gegevens te controleren en, waar nodig, te corrigeren.

In paragraaf 7.1 beschrijven we hoe fouten kunnen ontstaan. In paragraaf 7.2 gaan we dieper in op het controleren van de gegevens. En in paragraaf 7.3 beschrijven we technieken om de fouten te corrigeren.

7.1 Bronnen van fouten

Doordat u in uw peiling slechts een steekproef van personen benaderd, kunt u nooit exact de waarde van allerlei populatiekenmerken berekenen. Het blijft bij schattingen. Die schattingen kunnen afwijken van de werkelijkheid. Elke steekproef opnieuw is anders, en levert dus een andere schatting op, en dus ook een andere afwijking. U kunt uitrekenen hoe groot de afwijking maximaal kan zijn. Dat noemen we de *onzekerheidsmarge*. In het volgende hoofdstuk leggen we uit hoe u die onzekerheidsmarge berekent.

Deze *steekproeffout* heeft u zelf ingebouwd in uw onderzoek. U heeft de omvang van deze fout onder controle. Zo kunt u de onzekerheidsmarge verkleinen door een grotere steekproef te trekken. Er zijn echter ook bronnen van fouten die u niet onder controle heeft. We beschrijven drie van die foutenbronnen: onderdekking, non-respons en meetfouten.

Onderdekking doet zich voor als het steekproefkader niet alle objecten uit de populatie bevat. Die objecten kunnen nooit in de steekproef komen. Als die objecten systematisch afwijken van objecten die wel in het steekproefkader zitten, dan kunnen schattingen voor populatiekenmerken ook een afwijking hebben. Een voorbeeld van onderdekking is het gebruik van een telefoonboek als steekproefkader. Mensen die niet in het telefoonboek staan (zoals grote groepen jongeren met een mobieltje) zullen dus geen deel kunnen uitmaken van het onderzoek.

Non-respons is het verschijnsel dat u van de personen die u in de steekproef hebt getrokken, de gewenste informatie niet krijgt. Oorzaken van non-respons zijn geen contact (de persoon is niet thuis), weigering, en niet in staat (de persoon is ziek of spreekt en andere taal). Als de non-respondenten in uw peiling afwijken van de respondenten, dan loopt u een groot risico om verkeerde conclusies te trekken uit de peiling. Een bekend voorbeeld hiervan is dat respondenten een grotere geneigdheid hebben om te gaan stemmen dan non-respondenten. Als u dus op grond van de

respons in uw peiling de opkomst bij een verkiezing zou willen schatten, dan komt u te hoog uit. We behandelen non-respons uitgebreider in hoofdstuk 9.

Er kunnen ook problemen ontstaan bij het stellen van vragen. We spreken dan van *meetfouten*. Een meetfout doet zich voor als het antwoord dat de respondent op een vraag geeft, niet juist is. Het antwoord wijkt af van de werkelijkheid. Een meetfout kan op allerlei manieren ontstaan. Een bekend verschijnsel is *satisficing*. Daarbij doen de respondenten niet hun best om het juiste antwoord te geven, maar ze komen met een makkelijk antwoord dat ook wel redelijk lijkt. Ook een onduidelijke vraagstelling kan leiden tot misverstanden, en daardoor tot verkeerde antwoorden. Het stellen van irrelevante vragen (door het volgen van de verkeerde route) kan irritatie opwekken, met alle gevolgen van dien voor de beantwoording van de vragen. Bij mondeling enquêteren kunnen verstoringen optreden als gevolg van interacties tussen respondent, enquêteur en eventuele andere aanwezigen bij het vraaggesprek, en door de situatie (omgeving) waarin het vraaggesprek plaatsvindt. Factoren die hierbij een rol spelen zijn huidkleur, geslacht, leeftijd, opleiding en sociale klasse van enquêteur en/of respondent. Waar de onderwerpen gevoelig liggen, bestaat de mogelijkheid dat de respondent alleen sociaalwenselijke antwoorden geeft.

Problemen kunnen ook ontstaan bij het stellen van vragen die betrekking hebben op het verleden van de respondent. Ze kunnen gebeurtenissen vergeten zijn of op een verkeerd moment in de tijd plaatsen. Dit soort verschijnselen, die ook wel *geheugeneffecten* worden genoemd, doen zich bijvoorbeeld voor bij vragen over aankopen van goederen, bioscoopbezoek en raadplegen van een arts.

Voorbeeld 7.1. Vergeten van contacten met de huisarts

In 1983 heeft het CBS uitgebreid onderzoek gedaan naar geheugeneffecten in een gezondheidsenquête. In die enquête werd aan de respondenten gevraagd het aantal contacten met de huisarts in de laatste drie maanden te rapporteren.

Uit dit onderzoek bleek dat naarmate een contact langer was geleden, de kans groter was dat de respondent dit vergat te melden. Het percentage niet gerapporteerde contacten nam met ongeveer 4% per week toe. Over de hele periode van drie maanden werd ongeveer een kwart van de contacten vergeten.

7.2 Controle

U wilt dat uw peiling correcte conclusies oplevert? Dan zult u moeten proberen fouten in de gegevens op te sporen, en ontdekte fouten te corrigeren. De eerste stap is dus het controleren van de ingevulde formulieren. Het is daarbij handig om daarbij drie soorten controles te onderscheiden.

Waardecontroles

Elke vraag in de vragenlijst heeft een *domein*. Dat is de verzameling van toegestane antwoorden. Een waardecontrole stelt vast of het antwoord op een vraag inderdaad

in het domein van de vraag ligt. Is dat niet het geval, dan is er sprake van een *waardefout*.

Een voorbeeld is de vraag naar de leeftijd van de respondent waarbij het domein is gedefinieerd als alle gehele getallen tussen 18 en 120. Als dan een waarde van 199 wordt ingevuld, dan is dit een waardefout.

Waardefouten komen vooral voor bij gebruik van papieren vragenlijsten. Bij elektronische vragenlijsten staat de computerprogrammatuur meestal niet toe dat u antwoorden buiten het domein invoert.

Relatiecontroles

Bij relatiecontroles gaat het om de combinatie van antwoorden op verschillende vragen. Controles op waardefouten hoeven niet tot fouten te leiden, terwijl toch de combinatie onmogelijk is. De leeftijd van iemand zou best 15 jaar kunnen zijn, en ook kan iemand gehuwd zijn, maar toch is de combinatie van beide antwoorden (een gehuwde persoon van 15 jaar) uitgesloten.

Een relatiecontrole stelt vast of de antwoorden op de betrokken vragen een geldige combinatie opleveren. Geldige combinaties voor leeftijd en burgerlijke staat zijn

- Leeftijd onder de 15 en niet gehuwd.
- Leeftijd vanaf 16 en niet gehuwd.
- Leeftijd vanaf 16 en gehuwd.

Is er sprake van een ongeldige combinatie, dan is dat een *relatiefout*. Een probleem bij een relatiefout is dat niet op voorhand duidelijk is welke van de erbij betrokken vragen het probleem heeft veroorzaakt. En misschien waren beide antwoorden wel fout. Dat maakt correctie van dit type fout niet zo eenvoudig. Geeft een man op dat hij 10 jaar is en gehuwd, dan kan de leeftijd fout zijn, of de burgerlijke staat, of allebei.

Het handmatig uitvoeren van relatiecontroles op papieren vragenlijsten is lastig, zeker als er meer dan twee variabelen bij betrokken zijn, en de desbetreffende vragen niet vlak bij elkaar staan in de vragenlijst. Relatiecontroles zitten vaak ingebouwd in CAPI of CATI-programmatuur. Dit is een van de belangrijkste redenen dat computergestuurde vragenlijsten betere antwoorden opleveren. Het opnemen van relatiecontroles in vragenlijsten op het internet is een punt van discussie. Er zijn deskundigen die zeggen dat je respondenten niet moet lastig vallen met controles, omdat ze dan afhaken. Anderen zeggen dat je het juist wel moet doen, omdat er toch al veel meetfouten in dergelijke vragenlijsten zitten.

Routecontroles

Controles op routefouten zijn alleen zinvol als er in de vragenlijst sprongopdrachten zijn opgenomen. Dat zijn instructies die ervoor zorgen dat de respondenten alleen relevante vragen krijgen en dat irrelevante vragen worden overgeslagen. Zo zullen vragen over werkomstandigheden alleen zin hebben voor mensen die werk hebben. En vragen over zoeken naar werk zult u waarschijnlijk alleen willen stellen aan mensen die geen werk hebben. Als er fouten worden gemaakt bij het volgen van

sprongopdrachten, dan zal dat ertoe leiden dat respondenten verkeerde vragen beantwoorden.

Routefouten kunnen optreden bij het gebruik van papieren vragenlijsten. Niets verhindert de respondent om naar keuze vragen te beantwoorden of over te slaan. Het is daardoor vrij simpel om in het verkeerde deel van de vragenlijst terecht te komen. Bij CAPI en CATI dwingt de programmatuur meestal de juiste route door de vragenlijst af. De computer, en niet de respondent, bepaalt de volgende vraag op basis van de eerder gegeven antwoorden. Het is dan niet mogelijk om routefouten te maken.

Bij vragenlijsten op het internet heeft u de keuze om die al of niet de route door de vragenlijst af te dwingen. Door het afdwingen van de route vermijdt u routefouten, maar die dwang zou irritaties kunnen oproepen bij de respondenten. En dat zou kunnen leiden tot afbreken van het invullen van het formulier. U kunt ook besluiten de route niet af te dwingen, maar dan loop u het risico dat relevante vragen worden overgeslagen.

7.3 Correctie

Bij de controle van de gegevens zult u merken dat soms antwoorden ontbreken en soms antwoorden fout zijn. Met die foute antwoorden kunt u niets. Daarom is hier in feite ook sprake van ontbrekende antwoorden. Hoe dan ook, het komt erop neer dat een bestand heeft waarin juiste gegevens ontbreken. Dat maakt een goede analyse van de gegevens lastig.

Het zou heel goed kunnen dat gegevens selectief ontbreken. Dat betekent dat de wel beschikbare gegevens geen representatief beeld geven van de populatie waaruit ze zijn verkregen. Verder is het ook nog zo dat veel analysetechnieken niet in staat zijn om rekening te houden met ontbrekende gegevens. Ze eisen dat alle gegevens aanwezig zijn, of interpreteren codes voor ontbrekende waarden als 'echte' waarden.

De meest voor de hand liggende aanpak zou kunnen zijn opnieuw contact op te nemen met de respondenten, ze te confronteren met het probleem, en vervolgens te vragen om de correcte antwoorden. In de praktijk is dit feitelijk onmogelijk. Respondenten vinden het meestal al erg genoeg om één keer te worden lastig gevallen voor een peiling, laat staan twee keer. De kans is daarom vrij klein om opnieuw medewerking te krijgen, zeker als u met de mededeling komt dat de respondenten een fout heeft gemaakt. Bovendien is deze een aanpak tijdrovend en kostbaar is.

Een andere aanpak van het probleem is de formulieren met ontbrekende gegevens weg te gooien. Dan beperkt u zich tot de formulieren die volledig zijn gevuld. Deze aanpak gaat uit van de veronderstelling dat u de records met ontbrekende gegevens kunt opvatten als een aselechte steekproef (met gelijke kansen) uit het gehele bestand. Helaas is dit in de praktijk meestal niet het geval. Bepaalde soorten problemen treden vaak juist bij selecte groepen uit de populatie op. Bovendien gooit u op deze

manier wel heel veel informatie weg. Immers, één fout in een formulier is al voldoende om het helemaal weg te gooien.

Om het probleem van de ontbrekende waarden op te lossen, wordt heel vaak een imputatietechniek toegepast. Bij *imputatie* vervangt u de ontbrekende waarde door een *synthetische waarde*. Dit is dus niet een echt antwoord, maar een schatting van het echte antwoord. Een dergelijke schatting kunt u op allerlei manieren maken en dat leidt tot verschillende imputatietechnieken.

Een hele simpele imputatietechniek is *imputatie van het gemiddelde*. Stel dat enkele respondenten weigeren de vraag naar hun inkomen te beantwoorden. Dan kunt u de ontbrekende waarden vervangen door het gemiddelde van de inkomens die wel zijn opgegeven.

U kunt imputatie van het gemiddelde alleen toepassen voor een vraag als daarmee een kwantitatieve variabele meet, zodat het gemiddelde een zinvolle grootheid is. Als waarden selectief ontbreken (bijvoorbeeld alleen hoge inkomens), dan zal imputatie van het gemiddelde het probleem van de selectiviteit niet oplossen. Deze techniek zal een eventuele vertekening niet verminderen. Een vervelende eigenschap van imputatie van het gemiddelde is ook nog dat u de onzekerheidsmarges niet meer correct kunt uitrekenen. Als u de formules daarvoor uit hoofdstuk 8 toepast, dan leveren die een te kleine waarde op. U krijgt daardoor de indruk dat de schattingen heel nauwkeurig zijn, terwijl ze dat in werkelijkheid niet zijn,

Een tweede imputatietechniek is *aselecte imputatie*. U vervangt hierbij een ontbrekende waarde door een waarde die u heeft geloot uit de wel beschikbare antwoorden op de vraag. Dus als iemand weigert zijn inkomen op te geven, dan vult u een willekeurig inkomen in dat afkomstig is uit de wel opgegeven inkomens.

Aselecte imputatie kunt u toepassen voor zowel kwantitatieve als kwalitatieve variabelen. Ook deze imputatietechniek lost het probleem van het selectief ontbreken van waarden niet op. Wel is het zo dat de berekening van de onzekerheidsmarge een acceptabele waarde oplevert,

Een derde imputatietechniek is *donor-imputatie*. Het idee hierachter is dat u voor een respondent met een ontbrekend antwoord een andere respondent opzoekt die daar heel veel op lijkt en die vraag wel heeft beantwoord. Van die andere respondent neemt u dan het antwoord over. Stel dat u een respondent hebt met een ontbrekend inkomen. Dan zoekt u in uw bestand naar een andere respondent met hetzelfde geslacht, dezelfde leeftijd, dezelfde opleiding en dezelfde baan. Er vanuit gaande dat alle respondenten met deze karakteristieken ongeveer hetzelfde inkomen hebben, kunt u dan het inkomen van de donor gebruiken voort het ontbrekende inkomen.

Wat u in feite doet bij donor-imputatie is het voorspellen van het inkomen van iemand op basis van de antwoorden op de andere vragen. Het zal duidelijk zijn dat naarmate dit voorspellingsmodel beter werkt, de eventuele selectiviteit van de ontbrekende waarden beter gecorrigeerd wordt.

We hebben hier slechts drie imputatietechnieken in het kort genoemd. Er zijn er echter nog veel meer. Meer informatie over imputatie kunt u bijvoorbeeld vinden in Bethlehem, Cobben & Schouten (2011).

Voorbeeld 7.2. Imputatie van het inkomen

Aan de hand van een fictief voorbeeld laten we zien wat het effect van de verschillende imputatietechnieken kan zijn. De tabel hieronder toont de gegevens van 11 personen die hebben meegedaan aan een peiling. Er is gevraagd naar het maandinkomen, het opleidingsniveau en het aantal jaren werkervaring.

Eén van die personen heeft geweigerd zijn inkomen op te geven. Om het bestaan compleet te maken wilt u een geschikte synthetische waarde invullen. Welke imputatietechniek gebruikt u hiervoor? Een blik op de gegevens maakt al snel duidelijk dat personen met een lage opleiding minder verdienen dan personen met een hoge opleiding. Er is dus een relatie tussen opleiding en inkomen. Verder is het ook nog zo dat het inkomen geleidelijk stijgt met het aantal jaren werkervaring. Een geschikte imputatietechniek zal van deze verbanden gebruik maken.

Persoon	Inkomen	Opleiding	Werkervaring
1	€ 2041	Laag	1
2	€ 2110	Laag	2
3	€ 2142	Laag	3
4	€ 2201	Laag	4
5	€ 2247	Laag	5
6		Laag	6
7	€ 4099	Hoog	1
8	€ 4204	Hoog	2
9	€ 4298	Hoog	3
10	€ 4401	Hoog	4
11	€ 4497	Hoog	5

We lopen een aantal imputatietechnieken langs en we beginnen met imputatie van het gemiddelde. Dat zou betekenen dat u het gemiddelde van de 10 beschikbare inkomens invult. Dit gemiddelde is gelijk aan € 3224. Dit is duidelijk geen geschikte waarde, want de inkomens van mensen met een lage opleiding liggen net boven de 2000.

Aselecte imputatie kan een waarde net boven de € 2000 opleveren, maar ook een waarde boven de 4000. Dat zou dus een verkeerde waarde kunnen zijn.

U zou kunnen overwegen om imputatie van het gemiddelde of aselecte imputatie te beperken tot de groep mensen met lage inkomens. Dat levert wel een beter waarde op, maar u houdt dan geen rekening met het effect van werkervaring op het inkomen.

Voor donor-imputatie zoekt u een persoon op die lijkt op de imputeren persoon. Dat zou dan persoon 5 zijn, met dezelfde opleiding (laag) en maar één jaar verschil in werkervaring. U zou dan bij persoon 6 voor het inkomen € 2247 invullen. Ook dan neemt u het effect van werkervaring op inkomen onvoldoende mee.

Als u de gegevens wat nader bekijkt, dan zult u vaststellen dat bij de laag opgeleiden het inkomen gemiddeld € 50 stijgt per werkervaringjaar. Dat maakt het mogelijk het volgende model te gebruiken om het inkomen van laag opgeleiden te voorspellen:

$$\text{Inkomen} = 1997 + 50 \times \text{Werkervaring.}$$

Als u dit model toepast voor persoon 6, dan komt u uit op een inkomen van

$$1997 + 50 \times 6 = 2297.$$

Op basis van de beschikbare gegevens is dit de beste voorspelling van het inkomen van persoon 6.

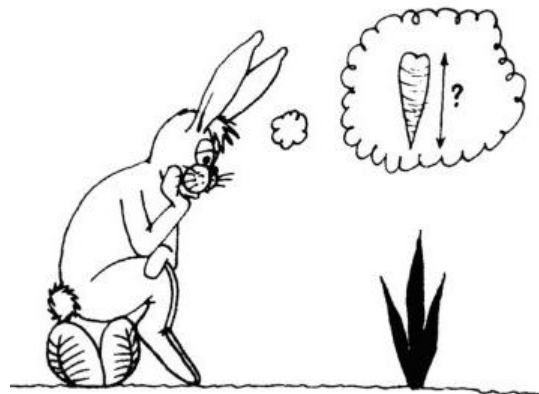
8. Schattingen maken

8.1 Schatters

De vragenlijst voor een peiling legt u voor aan alle personen in de steekproef. Als alles verder goed gaat, dan vullen die personen de vragenlijst allemaal keurig in. Zo krijgt u dus de waarden van de doelvariabelen en de hulpvariabelen. Uiteraard komen de waarden van de doelvariabele alleen beschikbaar voor de personen in de steekproef. Toch kunt u op basis van deze gegevens uitspraken doen over de populatie als geheel. Die uitspraken nemen de vorm aan van *schattingen* van populatiegrootheden.

Voor het berekenen van een schatting gebruikt u een *schatter*. Een schatter is een recept. Dit recept beschrijft welke berekeningen u moet uitvoeren om tot een schatting te komen. Het recept maakt ook duidelijk welke ingrediënten nodig zijn voor de berekening. Uiteraard zijn dat de steekproefgegevens. Soms is echter ook mogelijk om aanvullende informatie te gebruiken voor het berekenen van betere schattingen.

Figuur 8.1. Schatten



Een schatter is alleen bruikbaar als hij een schatting oplevert die dicht in de buurt ligt van de waarde die u wilt schatten. Daarom leggen we aan schatters twee eisen op:

- De schatter moet *zuiver* zijn. Stel eens dat u het trekken van de steekproef een groot aantal malen zou herhalen. Dat levert elke keer een andere steekproef op. Immers, het toeval bepaalt welke objecten worden geselecteerd. Dus levert de berekening van de schatting ook steeds weer een andere waarde op. Er is sprake van een *zuivere schatter* als het gemiddelde van alle mogelijke uitkomsten precies gelijk is aan de waarde die u wilt schatten. Of anders gezegd: herhaald trekken van een steekproef leidt niet tot een systematische onderschatting of overschatting. Gemiddeld genomen zal de schattingsprocedure de correcte waarde opleveren.

- De schatter moet *precies* zijn. Elke nieuwe steekproef levert een andere uitkomst voor de schatter op. Bij voorkeur moet de variatie in de uitkomsten zo klein mogelijk zijn. Alle mogelijke schattingen moeten zo dicht mogelijk bij elkaar in de buurt liggen.

Het begrip zuiverheid is gerelateerd aan het begrip validiteit. We noemen een meetinstrument *valide* als het meet wat we beogen te meten. Dus een valide meetinstrument leidt tot een zuivere schatter.

Het begrip precisie is gerelateerd aan het begrip betrouwbaarheid. Een meetinstrument is *betrouwbaar* als het bij herhaald gebruik (bij benadering) dezelfde schattingen oplevert. Dus een betrouwbaar meetinstrument leidt tot een precieze schatter.

Om de precisie van een schatter aan te geven, gebruiken we meestal een grootte die we de *variantie* noemen. De variantie kunt u opvatten als een soort gemiddelde afwijking van de mogelijke schattingen van de populatiewaarde. Is de waarde van de variantie klein, dan is er sprake van een precieze schatter. Grote waarden van de variantie duiden op een schatter met weinig precisie.

Merk op dat zowel zuiverheid als precisie belangrijk zijn. Een schatter die wel zuiver is, maar niet precies, kan toevallig een waarde opleveren die ver uit de buurt ligt van de werkelijke waarde. Een schatter die wel precies is, maar niet zuiver, levert systematisch verkeerde waarden op. Als de schatter zuiver en precies is, dan noemen we dat een *nauwkeurige schatter*. Voor een nauwkeurige schatter ligt de schatting met zeer grote waarschijnlijkheid dichtbij de te schatten waarde.

De keuze van de schatter, het recept voor de berekening van een schatting, in combinatie met de manier waarop u de steekproef trekt, bepaalt of de schatter zuiver is. De precisie van een schatter wordt voor een belangrijk deel bepaald door de omvang van de steekproef.

Voorbeeld 8.1. Betrouwbaarheid en validiteit van een weegschaal

We illustreren de begrippen betrouwbaarheid en validiteit aan de hand van een weegschaal. Stel dat u een onderzoek wilt doen naar overgewicht bij mensen. U gebruikt een weegschaal om het gewicht van een groot aantal mensen te meten.

Stel dat u een persoon die in werkelijkheid een wicht van 90 kilo heeft, vijf keer weegt. Als dan de weegschaal achtereenvolgens de waarden 80, 100, 90, 85 en 95 aanwijst, dan kunt u concluderen dat de weegschaal als meetinstrument niet erg betrouwbaar is. Er zit immers nogal wat variatie in de uitkomsten. De precisie is gering. De meting met de weegschaal is wel valide, want gemiddeld genomen wijst hij de juiste waarde aan. Hij heeft geen systematische afwijking.

Zouden alle uitkomsten geconcentreerd liggen om dat 100 kilo, dan is de weegschaal wel betrouwbaar maar niet valide. Zouden alle uitkomsten geconcentreerd liggen rond de 90, dan is de weegschaal betrouwbaar en valide.

In dit hoofdstuk zullen we schatters bespreken voor het schatten van een populatiepercentage en het populatiegemiddelde. In paragraaf 8.2 gaat het om schatters voor een enkelvoudige aselechte steekproef. Hierbij heeft elke persoon dezelfde kans om de in de steekproef te komen. Dit leidt tot vrij simpele recepten. Anders is het bij adressensteekproeven waarbij we één persoon per geselecteerd huishouden loten. Dan is er sprake van ongelijke kansen, en daarom krijgen de schatters ook een andere vorm. Deze schatters behandelen we in paragraaf 8.3. Als we weten hoe we de nauwkeurigheid van een schatter moeten berekenen, dan kunnen we ook laten zien hoe groot de omvang van een steekproef moet zijn om die nauwkeurigheid in de praktijk te realiseren. Dat doen we in paragraaf 8.4.

8.2 Schatters voor een enkelvoudige aselechte steekproef

Deze paragraaf behandelt schatters die toepasbaar zijn bij een enkelvoudige aselechte steekproef. Dit betekent dat iedereen in de doelpopulatie *dezelfde kans* moet hebben gehad om in de steekproef te komen. Bovendien moet de steekproef *zonder teruglegging* zijn getrokken.

8.2.1 Schatten van een populatiepercentage

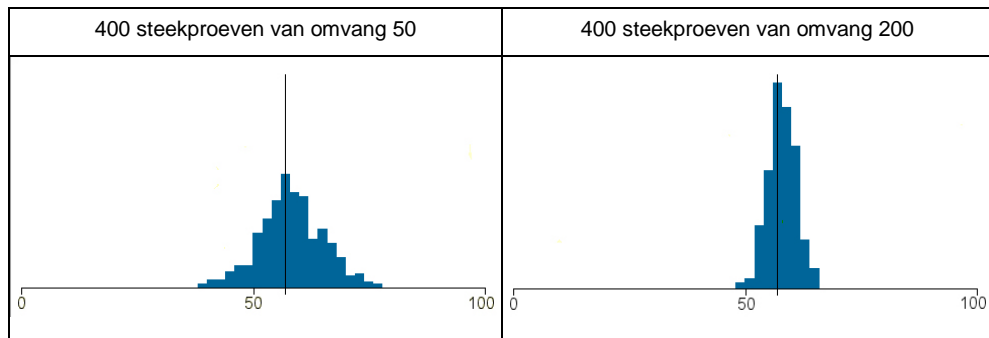
Eerst behandelen we het schatten van een populatiepercentage. Een voorbeeld hiervan is het percentage mensen in een gemeente dat naar de lokale omroep luistert. Bij een aselechte steekproef met gelijke kansen is heel vaak (maar niet altijd) het *analogieprincipe* van toepassing. Dit principe zegt dat als u een bepaalde grootheid in de populatie wilt schatten, u hiervoor de analoge grootheid in de steekproef kunt gebruiken. Het analogieprincipe is hier van toepassing voor het schatten van het populatiepercentage. Een goede schatter voor het populatiepercentage is het *steekproefpercentage*. Wilt u het percentage mensen schatten dat naar de lokale omroep luistert, dan neemt u daarvoor het percentage in de steekproef dat luistert.

Het steekproefpercentage is een zuivere schatter voor het populatiepercentage. Dat kunnen we wiskundig aantonen. Er is ook nog een andere manier om dat te laten zien, en dat is het uitvoeren van een simulatie. Eerst is met de computer een denkbeeldige doelpopulatie gemaakt. Die bestaat uit 15.000 personen. Daarvan luisteren 8.535 personen (56,9%) wel eens naar de lokale omroep. Vervolgens kunnen we de computer een groot aantal malen een enkelvoudige aselechte steekproef laten trekken. Voor elke steekproef berekenen we het percentage luisteraars. Zo krijgen we een groot aantal schattingen voor het percentage luisteraars in de populatie. Die schattingen kunnen we grafisch weergeven in de vorm van een histogram. In figuur 8.2 is dat gedaan. Elk blokje stelt een schatting voor. Leveren schattingen (afgerond) dezelfde waarde op, dan worden de blokjes op elkaar gestapeld.

De linker grafiek in figuur 8.2 is ontstaan door 400 steekproeven van omvang 50 te trekken. De verticale lijn in de grafiek geeft de populatiewaarde aan (56,9%). Alle blokjes liggen keurig om deze lijn verspreid. Het is ongeveer een 1-toppige, symmetrische figuur. De meeste schattingen liggen in de buurt van de populatie-

waarde: soms wat te laag, soms wat te hoog, maar gemiddeld goed. Er is hier sprake van een zuivere schatter.

Figuur 8.2. Simulatie van een enkelvoudige aselechte steekproef (percentage)



De rechter grafiek laat zien wat er gebeurt als u de omvang van de steekproef verhoogt van 50 naar 200. Het verschil met de linker grafiek is dat nu de schattingen veel dichter in de buurt van de populatiewaarde liggen. Er zit veel minder variatie in de mogelijke uitkomsten. Op basis van een steekproef van omvang 200 kunt u dus een veel preciezer schatting maken. Dit is een algemeen principe bij steekproeven: hoe groter de steekproef, des te preciezer de schatter.

U moet in de praktijk altijd aangeven hoe precies uw schattingen zijn. Alleen dan kunnen de gebruikers van de uitkomsten van uw peiling de conclusies op de juiste waarde schatten. De precisie geeft u aan met de onzekerheidsmarge of het betrouwbaarheidsinterval.

De *onzekerheidsmarge* beschrijft hoeveel uw schatting maximaal van de werkelijke waarde kan afwijken. De berekening van de onzekerheidsmarge gaat in de volgende stappen:

- (1) Bereken van de *variantie van de schatter*;
- (2) Gebruik deze variantie om de *standaardfout van de schatter* te berekenen. Dit is de wortel uit de variantie van de schatter;
- (3) Gebruik de standaardfout van de schatter om de *onzekerheidsmarge* te berekenen. Dit is standaardfout vermenigvuldigd met 1,96.

U kunt daarna ook nog het *betrouwbaarheidsinterval* uitrekenen. De ondergrens van dit interval krijgt u door de onzekerheidsmarge af te trekken van de schatting. De bovengrens krijgt u door de onzekerheidsmarge erbij op te tellen.

Er zijn geen specifieke populatiekenmerken voor kwalitatieve variabelen. Wat u wel kunt doen is het tellen van aantallen objecten in de verschillende categorieën. En die aantallen kunt u eventueel omwerken naar percentages. In feite hebt u dan de kwalitatieve variabele omgezet in een reeks indicatorvariabelen, waarbij er een indicatorvariabele is voor elke categorie.

Variantie van het steekproefpercentage

De variantie van een schatter geeft aan hoeveel de mogelijk uitkomsten van een schatter kunnen variëren. Is de variantie klein, dan is de schatter precies.

Laat P het te schatten populatiepercentage zijn, en p het percentage in de steekproef. Dan is de variantie van p gelijk aan

$$\text{Var}(p) = \left(\frac{1}{n} - \frac{1}{N} \right) \times \frac{N}{N-1} \times P \times (100 - P).$$

Hierin is N de omvang van de populatie, en n de omvang van de steekproef. Als de omvang van de populatie erg groot is, en de steekproefomvang veel kleiner dan de populatieomvang, dan kunt u bovenstaande formule vereenvoudigen tot:

$$\text{Var}(p) = \frac{P \times (100 - P)}{n}.$$

De omvang van de steekproef staat in de noemer van de formule. Dus de variantie is kleiner naarmate de steekproef groter is.

In de praktijk kunt u de variantie niet berekenen. Daar hebt u immers de waarde van het populatiepercentage P nodig, en die is nu juist onbekend. Wat u wel kunt doen, is de waarde van de variantie schatten op basis van je steekproefgegevens. Dat doet u door de waarde van P in de formules hierboven te vervangen door het steekproefpercentage p . Dit levert dan de formule

$$\text{var}(p) = \left(\frac{1}{n} - \frac{1}{N} \right) \times \frac{n}{n-1} \times p \times (100 - p)$$

op, of de vereenvoudigde vorm

$$\text{var}(p) = \frac{p \times (100 - p)}{n}.$$

We keren nog even terug naar het simulatie-experiment. Voor steekproeven van omvang 50 is de waarde van de variantie gelijk aan 49,0. Voor steekproeven van omvang 200 daalt de variantie naar 12,3. Dat is vier maal zo klein. Dit bevestigt de al eerder beschreven regel dat de precisie van een schatter toeneemt als de omvang van de steekproef groter wordt.

De variantie is een maat voor de precisie van een schatter. In de praktijk werkt deze maat echter niet zo handig. Wat zegt het nu dat de variantie gelijk is aan 12,3? Er is meer behoefte een maat die zegt hoe ver een schatting maximaal van de populatiewaarde kan afwijken. Deze maat is de *onzekerheidsmarge*. U kunt de onzekerheidsmarge gebruiken om een betrouwbaarheidsinterval uit te rekenen.

Het *betrouwbaarheidsinterval* heeft een ondergrens en een bovengrens. Die twee grenzen berekent u op basis van de in de steekproef beschikbaar gekomen gegevens. Die grenzen (schatting min onzekerheidsmarge en schatting plus onzekerheidsmarge) zijn zo bepaald dat de kans dat het interval de (onbekende) populatiewaarde

omvat, minstens gelijk is aan een van te voren vastgestelde grote kans. Die grote kans noemen we de *betrouwbaarheid*. Heel vaak hanteren we een betrouwbaarheid van 95%. In dit geval gaat het dan om het 95%-betrouwbaarheidsinterval. Door de standaardfout van de schatter met 1,96 te vermenigvuldigen krijgt u een 95%-betrouwbaarheidsinterval.

U kunt ook kiezen voor een grotere betrouwbaarheid. U krijgt, bijvoorbeeld, een 99%-betrouwbaarheidsinterval door de standaardfout van de schatter te vermenigvuldigen met 2,58. Dit leidt wel tot een groter betrouwbaarheidsinterval. Dit is de prijs die u betaald voor een grotere betrouwbaarheid: een grotere onzekerheidsmarge.

U kunt het 95%-betrouwbaarheidsinterval ook als volgt interpreteren: Als u het trekken van de steekproef en het vervolgens berekenen van de schatting een groot aantal malen zou herhalen, dan bevat het betrouwbaarheidsinterval in gemiddeld 95 van de 100 gevallen de te schatten waarde.

Merk op dat we op basis van steekproeven nooit uitspraken met absolute zekerheid kunnen doen. Er is altijd een element van onzekerheid. Dit wordt veroorzaakt door het loten van de steekproef. Daardoor is het mogelijk (maar de kans is heel klein) dat we per ongeluk een heel rare steekproef trekken. Het is bijvoorbeeld niet uitgesloten (maar het is zeer onwaarschijnlijk) dat in de populatie heel veel mensen naar de omroep luisteren, maar dat in de steekproef (per ongeluk) alleen maar mensen zitten die nooit naar de omroep luisteren.

Voor de berekening van het 95%-betrouwbaarheidsinterval moet u eerste de *standaardfout* bepalen. Die krijgt u door de wortel te trekken uit de variantie. Vervolgens moet u de *onzekerheidsmarge* uitrekenen. De waarde hiervan is gelijk aan de standaardfout vermenigvuldigd met 1,96. De *ondergrens* van het betrouwbaarheidsinterval krijgt u nu door de onzekerheidsmarge af te trekken van de schatting. En de *bovengrens* krijgt u door bij de schatting de marge op te tellen.

Betrouwbaarheidsinterval voor een percentage

De standaardfout van de schatting p voor het populatiepercentage P is gelijk aan

$$S(p) = \sqrt{\text{Var}(p)}$$

De marge M van het 95%-betrouwbaarheidsinterval is gelijk aan

$$M = 1,96 \times S(p) .$$

De ondergrens van het betrouwbaarheidsinterval is gelijk aan

$$p - M$$

en de bovengrens is gelijk aan

$$p + M .$$

Merk op dat u in de praktijk het betrouwbaarheidsinterval nooit exact kunt berekenen. Immers, daarvoor moet u de waarde van de variantie weten, maar die kunt u niet exact uitrekenen. Wel kunt u de variantie schatten op basis van de steekproef.

Daarmee kunt u vervolgens de standaardfout schatten. En daarmee kunt u weer schattingen voor de marge, en dus voor de onder- en bovengrens van het betrouwbaarheidsinterval berekenen. Zo krijgt u dus een geschat betrouwbaarheidsinterval.

Voorbeeld 8.2. Betrouwbaarheidsinterval voor het percentage luisteraars

Uit een doelpopulatie van 19.000 inwoners trekt u een enkelvoudige aselechte steekproef van 1.200 inwoners. 720 personen in de steekproef zeggen wel eens naar de lokale omroep te luisteren.

Het percentage luisteraars in de steekproef is gelijk aan $100 \times (720 / 1200) = 60\%$. De schatting voor het percentage luisteraars in de populatie is dus ook 60%.

De schatting voor de variantie van de schatting is gelijk aan $(1/1200 - 1/19000) \times (19000/18999) \times 60 \times 40 = 1,874$.

De schatting van de standaardfout krijgen we door het trekken van de wortel uit 1,874. Dat levert de waarde 1,369 op.

Als we de standaardfout vermenigvuldigen met 1,96, dan krijgen we de marge van het 95%-betrouwbaarheidsinterval. Dit levert (afgerond) de waarde 2,7 op.

De ondergrens van het betrouwbaarheidsinterval is gelijk aan $60 - 2,7 = 57,3$. De bovengrens van het betrouwbaarheidsinterval is gelijk aan $60 + 2,7 = 62,7$.

We kunnen dus met een waarschijnlijkheid van 95% stellen dat het percentage luisteraars in de populatie zal liggen tussen 57,3% en de 62,7%.

Merk op dat gebruik van de vereenvoudigde variantieformule een waarde van 2,000 zou hebben opgeleverd voor de geschatte variantie. Dit zou hebben geleid tot een betrouwbaarheidsinterval met een ondergrens van 57,2 en een bovengrens van 62,8.

8.2.2 Schatten van een populatiegemiddelde

Een voorbeeld van een populatiegemiddelde is het gemiddelde aantal uren per week dat inwoners van een gemeente naar de lokale omroep luisteren. Bij een enkelvoudige aselechte steekproef is ook hier weer het analogieprincipe van toepassing: het *steekproefgemiddelde* is een goede schatter voor het populatiegemiddelde. Wilt u het gemiddeld aantal uren schatten dat men naar de lokale omroep luistert, dan neemt u daarvoor het gemiddelde aantal uren in de steekproef dat men luistert.

Het steekproefgemiddelde

Het steekproefgemiddelde van een doelvariabele Y is gelijk aan

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{y_1 + y_2 + \dots + y_n}{n}$$

Hierin is n de omvang van de steekproef. De n waarden die in de steekproef beschikbaar komen, geven we aan met

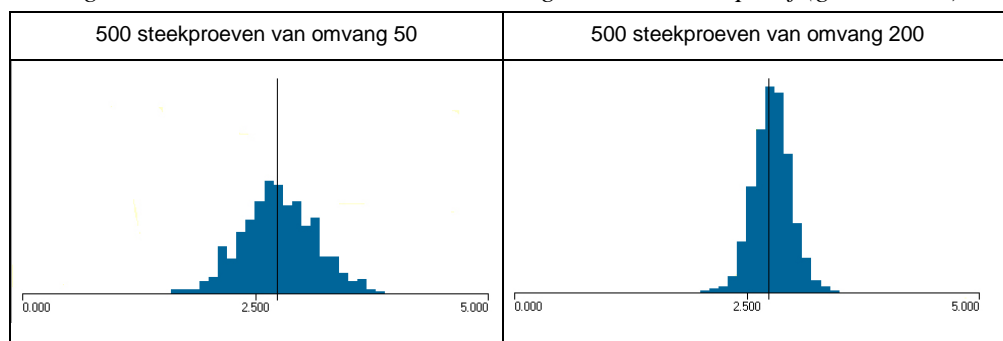
$$y_1, y_2, \dots, y_n.$$

Merk op dat we alle grootheden die betrekking hebben op de steekproef weergegeven met kleine letters.

Het steekproefgemiddelde is een zuivere schatter voor het populatiegemiddelde. Dat kunnen we wiskundig aantonen. Maar we laten het hier weer zien aan de hand van een simulatie. Eerst is met de computer een denkbeeldige doelpopulatie gemaakt. Die bestaat uit 15,000 personen. Bij de 8,535 personen die luisteren naar de omroep is aangegeven hoeveel uur ze de afgelopen week hebben geluisterd. Voor de mensen die niet hebben geluisterd, staat die waarde uiteraard op 0. Vervolgens kunnen we de computer een groot aantal malen een enkelvoudige aselechte steekproef laten trekken. Voor elke steekproef bepalen we het gemiddelde aantal luisteruren. Zo krijgen we een groot aantal schattingen. Die schattingen kun we grafisch weergeven in de vorm van een histogram. In figuur 8.3 hebben we dat gedaan. Elk blokje stelt een schatting voor. Leveren schattingen (afgerond) dezelfde waarde op, dan worden de blokjes op elkaar gestapeld.

De linker grafiek in figuur 8.3 is ontstaan door 500 steekproeven van omvang 50 te trekken. De verticale lijn in de grafiek geeft het populatiegemiddelde aan (2,7 uur). Alle blokjes liggen keurig om deze lijn verspreid. Het is ongeveer een 1-toppige, symmetrische figuur. De meeste schattingen liggen in de buurt van de populatiewaarde: soms te laag, soms te hoog, maar gemiddeld goed. Er is hier dus sprake van een zuivere schatter.

Figuur 8.3. Simulatie van een enkelvoudige aselechte steekproef (gemiddelde)



De rechter grafiek laat zien wat er gebeurt als we steekproeven van omvang 200 trekken. Het verschil met de linker grafiek is dat nu dat de schattingen veel dichter in de buurt van de populatiewaarde liggen. Er zit veel minder variatie in de

mogelijke uitkomsten van de schatter. Op basis van een steekproef van omvang 200 kunt u dus een veel nauwkeuriger schatting maken.

De precisie van het steekproefgemiddelde als schatter voor het populatiegemiddelde geven we ook weer aan met de onzekerheidsmarge of het betrouwbaarheidsinterval. Eerst berekenen we de variantie van het steekproefgemiddelde. Door daaruit de wortel te trekken, krijgen we de standaardfout van het steekproefgemiddelde. Vervolgens bepalen we de onzekerheidsmarge door de standaardfout te vermenigvuldigen met 1,96. Ten slotte vinden we de ondergrens van het 95%-betrouwbaarheidsinterval door de onzekerheidsmarge van het steekproefgemiddelde af te trekken. En voor de bovengrens tellen we de onzekerheidsmarge op bij het steekproefgemiddelde.

Variantie van het steekproefgemiddelde

Laat \bar{Y} het te schatten populatiegemiddelde voorstellen, en \bar{y} het gemiddelde in de steekproef. Dan is de variantie van het steekproefgemiddelde gelijk aan

$$\text{Var}(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N} \right) \times S^2.$$

Hierin is N de omvang van de populatie en n de omvang van de steekproef. De aangepast populatievariantie S^2 is al in hoofdstuk 2 (paragraaf 2.4) ingevoerd:

$$S^2 = \frac{1}{N-1} \sum_{k=1}^N (Y_k - \bar{Y})^2.$$

Als de omvang van de populatie erg groot is, en de steekproefomvang veel kleiner dan de populatieomvang, dan kunt u een wat simpeler formule hanteren voor de variantie:

$$\text{Var}(\bar{y}) = \frac{S^2}{n}.$$

Merk op dat de steekproefomvang n in de noemer van deze uitdrukking staat. Dus als de omvang van de steekproef toeneemt, dan wordt de variantie kleiner, en dus de schatting preciezer.

U kunt de variantie in de praktijk niet exact berekenen. Daarvoor zijn immers alle waarden van de doelvariabele nodig, en die zijn nu juist onbekend. De oplossing is het schatten van de variantie op basis van de steekproef. Daarvoor vervangt u S^2 door de overeenkomstige grootte voor de steekproef:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Dit leidt dan tot een schatting voor de variantie:

$$\text{var}(\bar{y}) = \frac{s^2}{n}.$$

We keren nog even terug naar ons simulatie-experiment. Voor steekproeven van omvang 50 is de waarde van de variantie van het steekproefgemiddelde gelijk aan 0,160. Voor steekproeven van omvang 200 daalt de variantie naar 0,040. Dat is vier maal zo klein. Een vier maal zo grote steekproef leidt dus tot een vier maal zo kleine variantie.

Betrouwbaarheidsinterval voor een gemiddelde

De standaardfout van het steekproefgemiddelde is gelijk aan

$$S(\bar{y}) = \sqrt{\text{Var}(\bar{y})}.$$

De onzekerheidsmarge van het 95%-betrouwbaarheidsinterval is gelijk aan

$$M = 1,96 \times S(\bar{y})$$

De ondergrens van het betrouwbaarheidsinterval is gelijk aan

$$p - M$$

en de bovengrens is gelijk aan

$$p + M.$$

In de praktijk kunt u het betrouwbaarheidsinterval niet exact berekenen, omdat u de variantie niet exact kunt berekenen. Daarvoor zijn immers alle waarden van de doelvariabele nodig, en die hebt u niet. Wel kunt u de variantie schatten op basis van de steekproef. Vervolgens kunt u een schatting van de standaardfout bepalen. En daarmee kunt u weer schattingen voor de onzekerheidsmarge, en dus voor de onder- en bovengrens van het betrouwbaarheidsinterval berekenen. Zo krijgt u dus een geschat betrouwbaarheidsinterval.

Voorbeeld 8.3. Betrouwbaarheidsinterval voor de gemiddelde luisterduur

Uit een doelpopulatie van omvang 15.000 inwoners trekken we een enkelvoudige aselecte steekproef van 20 inwoners. Aan de personen in de steekproef vragen we hoeveel uur ze afgelopen week naar de lokale omroep hebben geluisterd. De gegevens staan in de tweede kolom van de tabel hieronder.

Het steekproefgemiddelde is gelijk aan de som van de waarden in de tweede kolom (56,40) gedeeld door 20. De uitkomst is 2,82 (uur).

Voor het berekenen van de variantie van de schatter moet u eerst de steekproefvariantie s^2 uitrekenen. Hiervoor trekt u van elke waarde het gemiddelde (2,82) af. Dat leidt tot kolom 3. Vervolgens kwadrateert u deze waarden, wat leidt tot kolom 4. Nu kunt u de steekproefvariantie s^2 bepalen door kolom 4 op te tellen, en de uitkomst te delen door $n - 1 = 19$. Dit levert de waarde 6,44 op.

	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	0,00	-2,82	7,95
2	3,40	0,58	0,34
3	4,60	1,78	3,17
4	4,10	1,28	1,64
5	3,90	1,08	1,17
6	0,00	-2,82	7,95
7	3,40	0,58	0,34
8	7,30	4,48	20,07
9	0,00	-2,82	7,95
10	0,00	-2,82	7,95
11	3,80	0,98	0,96
12	0,00	-2,82	7,95
13	0,00	-2,82	7,95
14	4,20	1,38	1,90
15	5,50	2,68	7,18
16	4,40	1,58	2,50
17	6,40	3,58	12,82
18	5,40	2,58	6,66
19	0,00	-2,82	7,95
20	0,00	-2,82	7,95

De variantie van de schatter krijgt u nu door de steekproefvariantie te vermenigvuldigen met $(1/n - 1/N) = (1/20 - 1/15000)$. Dit levert de waarde 0,32 op. De schatting van de standaardfout is gelijk aan de wortel uit 0,32. Dat levert de waarde 0,57 op.

De onzekerheidsmarge van het 95%-betrouwbaarheidsinterval krijgt u door de standaardfout te vermenigvuldigen met 1,96. Dit levert (afgerond) de waarde 1,11 op.

De ondergrens van het betrouwbaarheidsinterval is gelijk aan $2,82 - 1,11 = 1,71$. De bovengrens van het betrouwbaarheidsinterval is gelijk aan $2,82 + 1,11 = 3,93$.

Met een waarschijnlijkheid van 95% kunt u dus stellen dat het gemiddelde aantal uur dat men per week luistert naar de lokale omroep zal liggen tussen 1,71 uur en de 3,93 uur.

Er is hier toch wel sprake is van een ruime marge. Die wordt veroorzaakt door de kleine steekproefomvang van 20.

8.3 Schatters voor een tweetrapssteekproef

In hoofdstuk 4 hebben we al uitgelegd dat bij het trekken van personen via een adressensteekproef de personen niet dezelfde kans hebben om in de steekproef te komen. De trekkingskans wordt mede bepaald door het aantal personen dat op het adres woont (en tot de doelpopulatie behoort). Voor dergelijke steekproeven kunnen we geen schattingen maken met de aanpak in de vorige paragraaf. Dat zou tot onjuiste (onzuivere) schattingen leiden. Om toch correcte schattingen te kunnen maken, moet we de theorie voor steekproeven met ongelijke kansen toepassen. In deze paragraaf leggen we uit hoe dit in zijn werk gaat.

We gaan uit van een onderzoeksopzet waarbij we eerst adressen loten met een enkelvoudige aselecte steekproef. Vervolgens loten we op elk geselecteerd adres één persoon uit de daar wonende personen (voor zover ze behoren tot de doelpopulatie).

De trekkingskansen voor adressen zijn wel gelijk, maar personen in grote huishoudens hebben een kleinere kans dan personen in kleine huishoudens. Bij het maken van schattingen op grond van een dergelijke tweetrapssteekproef moeten we voor deze ongelijke trekkingskansen corrigeren. Dat kan alleen als we de trekkingskansen weten. Daarvoor is het noodzakelijk dat we het aantal personen (voor zover behorend tot de doelpopulatie) op elk geselecteerd adres registreren.

Voorbeeld 8.4. Een adressensteekproef voor een luisteronderzoek

In een luisteronderzoek is op 209 adressen iemand geïnterviewd. Het totale aantal adressen in de gemeente was gelijk aan 9.590. De trekkingskans van elk adres is dus gelijk aan $209 / 9590$.

Verder is de trekkingskans van een persoon op een geselecteerd adres afhankelijk van het aantal daar wonende personen van 12 jaar en ouder. Als we dit aantal aangeven met A , dan is die trekkingskans gelijk aan:

$$\frac{1}{A}.$$

De totale kans voor een persoon om in de steekproef te komen, krijgen we door beide bovenstaande kansen met elkaar te vermenigvuldigen. Dit geeft als trekkingskans:

$$\frac{209}{A \times 9590}.$$

Niet elke persoon in de doelpopulatie heeft dus dezelfde kans om in de steekproef te komen. Een persoon in een 1-persoons huishouden heeft een kans van $209 / 9590 = 0,022$. De kans voor een persoon in een 2-persoons huishouden is twee maal zo klein: $209 / 19180 = 0,011$. Het gevolg is dat personen in kleine huishoudens oververtegenwoordigd zijn in de steekproef en personen uit grote huishoudens zijn ondervertegenwoordigd.

Als de trekkingskansen ongelijk zijn, dan is het in de vorige paragraaf genoemde *analogieprincipe* niet meer van toepassing. Het steekproefpercentage en het steekproefgemiddelde zijn geen zuivere schatters meer voor populatiepercentage en populatiegemiddelde. We moeten andere schattingsprocedures gebruiken waarmee we corrigeren voor deze ongelijke kansen.

8.3.1 Schatten van een populatiepercentage

We beginnen met het schatten van een populatiepercentage. Een voorbeeld hiervan is het percentage mensen dat naar de lokale omroep luistert. Aan elke persoon in de steekproef kennen we een gewicht toe. Dit gewicht berekenen we als volgt:

$$\text{Gewicht} = \frac{\text{Aantal personen op adres} \times \text{Aantal adressen in doelpopulatie}}{\text{Omvang doelpopulatie}}.$$

Bij het aantal personen op het adres gaat het uiteraard alleen om de personen die tot de doelpopulatie behoren. Een speciaal geval is de situatie waarin er op elk adres evenveel personen wonen. Dan zijn alle gewichten gelijk aan 1. Er is dan dus geen sprake van een correctie. Dat hoeft ook niet, aangezien iedereen dezelfde kans heeft.

Voorbeeld 8.5. Berekening van de gewichten voor een luisteronderzoek

In een luisteronderzoek wordt op 209 adressen iemand geïnterviewd. Het totale aantal adressen in de doelpopulatie is gelijk aan 9.590. De omvang van de doelpopulatie (alle inwoners van 12 jaar en ouder) is 23.126. Toepassing van de formule voor de gewichten leidt dan tot onderstaande tabel:

Aantal personen op het adres	Gewicht van persoon in steekproef
1	0,415
2	0,829
3	1,244
4	1,659
5	2,073
6	2,488
...	...

Duidelijk is te zien dat een persoon in een 1-persoons huishouden een lager gewicht krijgt dan een persoon in een meer-persoons huishouden. Dat is logisch want personen uit 1-persoons huishoudens zijn oververtegenwoordigd in de steekproef. Om hiervoor te corrigeren moeten we ze minder zwaar meetellen in de schattingsprocedure.

Als we de gewichten hebben bepaald, dan kunnen we een zuivere schatter uitrekenen voor het percentage personen in de populatie met een bepaalde eigenschap. Die schatter is gedefinieerd als

$$\text{Schatting} = 100 \times \frac{\text{Som van de gewichten van de personen met die de eigenschap}}{\text{Omvang van de steekproef}}.$$

Merk op dat als alle personen in de steekproef hetzelfde gewicht zouden hebben (zoals in de situatie van de enkelvoudige aselecte steekproef uit de vorige paragraaf), deze schatting overgaat in het simpele steekproefpercentage.

Schatting van een percentage in een steekproef van adressen

Stel Y is een indicatorvariabele die aangeeft of een persoon in de doelpopulatie een bepaalde eigenschap wel (waarde = 1) of niet (waarde = 0) heeft. Het populatiepercentage P is dan gelijk aan 100 maal het populatiegemiddelde van Y .

De totale populatie van N personen is verdeeld over M adressen. De aantallen personen per adres worden genoteerd met

$$A_1, A_2, \dots, A_M.$$

Optellen van al die aantallen geeft dus de totale omvang van de populatie N .

Uit deze doelpopulatie loten we eerst n adressen. Vervolgens loten we op elke geselecteerd adres 1 persoon. De steekproef bestaat dus uit n personen. We geven de gemeten waarden van de doelvariabele voor die n personen aan met

$$y_1, y_2, \dots, y_n$$

(waarbij alleen de waarden 0 of 1 worden aangenomen). Het aantal personen op de geselecteerde adressen geven we aan met

$$a_1, a_2, \dots, a_n$$

Het gewicht w_i van persoon i is gelijk aan

$$w_i = a_i \frac{M}{N},$$

waarbij i loopt van 1 t/m n .

De schatting voor het populatiepercentage wordt nu

$$p_w = \frac{100}{n} \times \sum_{i=1}^n w_i y_i.$$

Voorbeeld 8.6. Schatting van een percentage in een steekproef van adressen

Uit een doelpopulatie van 20.000 inwoners, verdeeld over 7.000 adressen, hebben we een enkelvoudige aselechte steekproef van 20 adressen getrokken. Op elk getrokken adres loten we één persoon. Aan de personen in de steekproef vragen we of ze de afgelopen week naar de lokale omroep hebben geluisterd. De gegevens benodigd voor het maken van een schatting, staan in de tabel hieronder:

Adres	Aantal personen	Geselecteerde persoon luistert	Gewicht	Gewicht luisteraar
1	2	Ja	0,70	0,70
2	1	Nee	0,35	
3	4	Ja	1,40	1,40
4	3	Ja	1,05	1,05
5	1	Ja	0,35	0,35
6	6	Ja	2,10	2,10
7	2	Nee	0,70	
8	1	Nee	0,35	
9	1	Nee	0,35	
10	4	Ja	1,40	1,40
11	2	Nee	0,70	
12	2	Ja	0,70	0,70
13	3	Nee	1,05	
14	4	Ja	1,40	1,40
15	1	Ja	0,35	0,35
16	3	Nee	1,05	
17	2	Nee	0,70	
18	1	Nee	0,35	
19	3	Ja	1,05	1,05
20	1	Nee	0,35	
Totaal				10,50

De som van de gewichten van de luisteraars is gelijk aan 10,50. De schatting voor het percentage luisteraars in de populatie wordt nu:

$$100 \times \frac{10,50}{20} = 52,5\% .$$

Merk op dat het gewone steekproefpercentage in deze steekproef gelijk is aan 50% (10 van de 20 mensen luisteren). Het onterechte gebruik van deze schatter levert hier dus een te lage waarde op.

Ook bij steekproeven met ongelijke kansen geldt het principe dat een grotere steekproef leidt tot preciezere schattingen. En die precisie kunnen we weer kwantificeren door het achtereenvolgend uitrekenen van de variantie, standaardfout, onzekerheidsmarge en betrouwbaarheidsinterval. De formules hiervoor zijn echter anders dan die voor de enkelvoudige aselecte steekproef.

De variantie van de schatter voor een percentage

Laat p_w de schatter zijn voor het populatiepercentage zoals hierboven beschreven. Dan is de variantie van p_w gelijk aan

$$\text{Var} (p_w) = \frac{10000}{nN} \sum_{k=1}^N \frac{(w_k Y_k - \bar{Y})^2}{w_k} ,$$

waarbij we sommeren over de hele populatie. Deze variantie kunt u in de praktijk niet berekenen. Daarvoor zijn alle waarden van de doelvariabele nodig, en die zijn nu juist onbekend. Wat u wel kunt doen is de waarde van de variantie schatten op basis van de steekproefgegevens. Dat doet u met de volgende formule:

$$\text{var}(p_w) = \frac{10000}{n(n-1)} \sum_{i=1}^n (w_i y_i - \bar{y})^2 ,$$

waarbij we sommeren over de steekproef.

De berekening van de onzekerheidsmarge en het 95%-betrouwbaarheidsinterval gaat op exact dezelfde manier als in paragraaf 8.2:

- (1) Bereken of schat de *standaardfout*. Die is gelijk aan de wortel uit de (geschatte) variantie van de schatter.
- (2) Bereken de *onzekerheidsmarge*. Die is gelijk aan de standaardfout vermenigvuldigd met 1,96.
- (3) Bereken de *ondergrens* van het betrouwbaarheidsinterval. Die krijgt u door de onzekerheidsmarge af te trekken van de schatting.
- (4) Bereken de *bovengrens* van het betrouwbaarheidsinterval. Die krijgt u door de onzekerheidsmarge op te tellen bij de schatting.

Voorbeeld 8.7. Betrouwbaarheidsinterval voor een percentage

Uit een doelpopulatie van 20.000 inwoners, verdeeld over 7.000 adressen, is een enkelvoudige aselecte steekproef van 20 adressen getrokken. Op elk getrokken adres is één persoon geloot. Aan deze personen is gevraagd of ze de afgelopen week naar de lokale omroep hebben geluisterd. De gegevens staan in de tabel:

Adres	a_i	y_i	$w_i = a_i \frac{M}{N}$	$w_i \times y_i$	$(w_i \times y_i - \bar{y})$	$(w_i \times y_i - \bar{y})^2$
1	2	1	0,70	0,70	0,175	0,030625
2	1	0	0,35	0,00	-0,525	0,275625
3	4	1	1,40	1,40	0,875	0,765625
4	3	1	1,05	1,05	0,525	0,275625
5	1	1	0,35	0,35	-0,175	0,030625
6	6	1	2,10	2,10	1,575	2,480625
7	2	0	0,70	0,00	-0,525	0,275625
8	1	0	0,35	0,00	-0,525	0,275625
9	1	0	0,35	0,00	-0,525	0,275625
10	4	1	1,40	1,40	0,875	0,765625
11	2	0	0,70	0,00	-0,525	0,275625
12	2	1	0,70	0,70	0,175	0,030625
13	3	0	1,05	0,00	-0,525	0,275625
14	4	1	1,40	1,40	0,875	0,765625
15	1	1	0,35	0,35	-0,175	0,030625
16	3	0	1,05	0,00	-0,525	0,275625
17	2	0	0,70	0,00	-0,525	0,275625
18	1	0	0,35	0,00	-0,525	0,275625
19	3	1	1,05	1,05	0,525	0,275625
20	1	0	0,35	0,00	-0,525	0,275625
Som				10,50	0,000	8,2075
Gemiddeld				0,525		

De variantie kunnen we nu uitrekenen door het totaal van de laatste kolom in te vullen in de formule voor de geschatte variantie. We krijgen dan

$$\text{var}(p) = \frac{10000}{20 \times 19} \times 8,2075 = 215,987 .$$

De standaardfout krijgen we door het trekken van de wortel hieruit:

$$s(p) = \sqrt{\text{var}(p)} = \sqrt{215,987} = 14,700 .$$

De onzekerheidsmarge van het 95%-betrouwbaarheidsinterval is gelijk aan

$$M = 1,96 \times s(p) = 1,96 \times 14,700 = 28,805 .$$

De ondergrens van het betrouwbaarheidsinterval is nu

$$p - M = 52,50 - 28,805 = 23,695$$

en de bovengrens is

$$p + M = 52,50 + 28,805 = 81,305 .$$

We kunnen dus met 95% zekerheid stellen dat het percentage luisteraars zal liggen tussen 23,7% en 81,3%. Dit is een zeer breed interval waar je niet zoveel mee kunt. De oorzaak hiervan is de zeer gering steekproefomvang van 20 elementen. Voor een nauwkeuriger schatting is een veel grotere steekproef nodig.

8.3.2 Schatten van een populatiegemiddelde

De procedure voor het schatten van een populatiegemiddelde in een adressen-steekproef is in globale termen dezelfde als die voor het schatten van een populatiepercentage: eerst rekenen we voor elk persoon in de steekproef een gewicht uit. Vervolgens nemen we die gewichten mee in het berekenen van een schatting. Een voorbeeld van een populatiegemiddelde is het gemiddelde aantal uren per week dat men naar de lokale omroep luistert.

Aan elke persoon in de steekproef kennen we een gewicht toe. Het is hetzelfde gewicht als bij het schatten van een populatiepercentage. Dit gewicht bepalen we met de formule

$$\text{Gewicht} = \frac{\text{Aantal personen op adres} \times \text{Aantal adressen in doelpopulatie}}{\text{Omvang doelpopulatie}}.$$

Bij het aantal personen op het adres gaat het uiteraard alleen om de personen die tot de doelpopulatie behoren. In de speciale situatie waarin er op elk adres evenveel personen zouden wonen, krijgt elk gewicht de waarde 1.

Voorbeeld 8.5 in paragraaf 8.3.1 toont hoe de berekening van de gewichten voor een luisteronderzoek verloopt.

Als de gewichten zijn bepaald, kunnen we een zuivere schatter uitrekenen voor het gemiddelde van een doelvariabele in de populatie. Die schatter is gedefinieerd als

$$\text{Schatting} = \frac{\text{Gewogen som in de steekproef van de waarden van de doelvariabele}}{\text{Omvang van de steekproef}}$$

Merk op dat als alle personen in de steekproef hetzelfde gewicht zouden hebben (zoals dat in feite het geval is in de situatie van de enkelvoudige aselecte steekproef in paragraaf 8.2), deze schatting overgaat in het simpele steekproefgemiddelde.

Schatting van een gemiddelde in een steekproef van adressen

Laat Y de doelvariabele van het onderzoek zijn. Het gaat dan om het schatten van het populatiegemiddelde van Y .

De totale populatie van N personen is verdeeld over M adressen. De aantallen personen per adres worden genoteerd met

$$A_1, A_2, \dots, A_M.$$

Optellen van al die aantallen geeft dus de totale omvang van de populatie N .

Uit deze doelpopulatie loten we eerst n adressen. Vervolgens loten we op elke geselecteerd adres 1 persoon. De steekproef bestaat dus uit n personen. We geven de gemeten waarden van de doelvariabele voor die n personen aan met

$$y_1, y_2, \dots, y_n.$$

Het aantal personen op de geselecteerde adressen geven we aan met

$$a_1, a_2, \dots, a_n.$$

Het gewicht w_i van persoon i is gelijk aan

$$w_i = a_i \frac{M}{N},$$

waarbij i loopt van 1 t/m n .

De schatting voor het populatiepercentage wordt nu

$$\bar{y}_w = \frac{1}{n} \times \sum_{i=1}^n w_i y_i.$$

Voorbeeld 8.8. Schatting van een gemiddelde in een steekproef van adressen

Uit een doelpopulatie van 20.000 inwoners, verdeeld over 7.000 adressen, trekken we een enkelvoudige aselecte steekproef van 20 adressen. Op elk getrokken adres loten we één persoon. Aan de personen in de steekproef vragen we hoeveel uur ze de afgelopen week naar de lokale omroep hebben geluisterd. De gegevens benodigd voor het maken van een schatting, staan in de tabel hieronder:

Adres	Aantal personen	Aantal uren geluisterd	Gewicht	Gewicht x Uren
1	2	0,00	0,70	0,000
2	1	3,40	0,35	1,190
3	4	4,60	1,40	6,440
4	3	4,10	1,05	4,305
5	1	3,90	0,35	1,365
6	6	0,00	2,10	0,000
7	2	3,40	0,70	2,380
8	1	7,30	0,35	2,555
9	1	0,00	0,35	0,000
10	4	0,00	1,40	0,000
11	2	3,80	0,70	2,660
12	2	0,00	0,70	0,000
13	3	0,00	1,05	0,000
14	4	4,20	1,40	5,880
15	1	5,50	0,35	1,925
16	3	4,40	1,05	4,620
17	2	6,40	0,70	4,480
18	1	5,40	0,35	1,890
19	3	0,00	1,05	0,000
20	1	0,00	0,35	0,000
Totaal				39,690

De gewogen som is hier gelijk aan 39,690. De schatting voor het gemiddelde aantal luisteruren in de populatie is nu:

$$\frac{39,690}{20} = 1,985.$$

Merk op dat het gewone steekproefpercentage in deze steekproef gelijk is aan 2,820 uur. Het onterechte gebruik van deze schatter levert hier dus een te hoge waarde voor de schatting van het populatiegemiddelde op.

Ook hier geldt weer de regel dat een grotere steekproef leidt tot preciezer schattingen. En die precisie kunnen we kwantificeren door het achtereenvolgend uitrekenen (of schatten) van de variantie, standaardfout, onzekerheidsmarge en betrouwbaarheidsinterval. De formules voor de variantie en variantieschatting zijn echter anders dan die voor de enkelvoudige aselecte steekproef.

De variantie van de schatter voor een gemiddelde

Laat \bar{y}_w de schatter zijn voor het populatiegemiddelde \bar{Y} zoals hierboven beschreven. Dan is de variantie van dit gewogen steekproefgemiddelde gelijk aan

$$\text{Var}(\bar{y}_w) = \frac{1}{nN} \sum_{k=1}^N \frac{(w_k Y_k - \bar{Y})^2}{w_k},$$

waarbij de index k loopt over alle objecten in de populatie. Deze variantie kunt u in de praktijk nooit berekenen. Daarvoor zijn alle waarden van de doelvariabele nodig, en die zijn nu juist onbekend. Wat u wel kunt doen is de waarde van de variantie schatten op basis van de steekproefgegevens. Dat doet u met de volgende formule:

$$\text{var}(\bar{y}_w) = \frac{1}{n(n-1)} \sum_{i=1}^n (w_i y_i - \bar{y}_w)^2,$$

waarbij de index i loopt over alle objecten in de steekproef.

Voor de berekening van het 95%-betrouwbaarheidsinterval moet u eerst de *standaardfout* bepalen. Die krijgt u door de wortel te trekken uit de (geschatte) variantie. Vervolgens berekent u de *onzekerheidsmarge*. De waarde hiervan is gelijk aan de standaardfout vermenigvuldigd met 1,96. De *ondergrens* van het betrouwbaarheidsinterval krijgt u nu door de onzekerheidsmarge af te trekken van de schatting. En de *bovengrens* krijgt u door bij de schatting de onzekerheidsmarge op te tellen.

Voorbeeld 8.9. Betrouwbaarheidsinterval voor een gemiddelde

Uit een doelpopulatie van 20.000 inwoners, verdeeld over 7.000 adressen, trekken we een enkelvoudige aselecte steekproef van 20 adressen. Op elk getrokken adres loten we één persoon.

Aan de geselecteerde personen is gevraagd hoeveel uur ze de afgelopen week naar de lokale omroep hebben geluisterd. De gegevens die nodig zijn voor het berekenen van een betrouwbaarheidsinterval, staan hieronder:

Adres	a_i	y_i	$w_i = a_i \frac{M}{N}$	$w_i \times y_i$	$(w_i \times y_i - \bar{y})$	$(w_i \times y_i - \bar{y})^2$
1	2	0	0,70	0,000	-1,985	3,938
2	1	3,4	0,35	1,190	-0,795	0,631
3	4	4,6	1,40	6,440	4,456	19,851
4	3	4,1	1,05	4,305	2,321	5,385
5	1	3,9	0,35	1,365	-0,620	0,384
6	6	0	2,10	0,000	-1,985	3,938
7	2	3,4	0,70	2,380	0,396	0,156
8	1	7,3	0,35	2,555	0,571	0,325
9	1	0	0,35	0,000	-1,985	3,938
10	4	0	1,40	0,000	-1,985	3,938
11	2	3,8	0,70	2,660	0,676	0,456
12	2	0	0,70	0,000	-1,985	3,938
13	3	0	1,05	0,000	-1,985	3,938
14	4	4,2	1,40	5,880	3,896	15,175
15	1	5,5	0,35	1,925	-0,060	0,004
16	3	4,4	1,05	4,620	2,636	6,946
17	2	6,4	0,70	4,480	2,496	6,228
18	1	5,4	0,35	1,890	-0,095	0,009
19	3	0	1,05	0,000	-1,985	3,938
20	1	0	0,35	0,000	-1,985	3,938
Som				39,690	0,000	87,056
Gemiddeld				1,985		

De variantie kunnen we nu berekenen door het totaal van de laatste kolom in te vullen in de formule voor de geschatte variantie. We krijgen dan

$$\text{var}(\bar{y}_w) = \frac{1}{20 \times 19} \times 87,056 = 0,229 .$$

De standaardfout krijgen we door het trekken van de wortel hieruit:

$$s(\bar{y}_w) = \sqrt{\text{var}(\bar{y}_w)} = \sqrt{0,229} = 0,479 .$$

De onzekerheidsmarge van het 95%-betrouwbaarheidsinterval is gelijk aan

$$M = 1,96 \times s(\bar{y}_w) = 1,96 \times 0,479 = 0,938 .$$

De ondergrens van het betrouwbaarheidsinterval is nu

$$p - M = 1,985 - 0,938 = 1,046$$

en de bovengrens is

$$p + M = 1,985 + 0,938 = 2,923 .$$

We kunnen dus met 95% zekerheid stellen dat het gemiddelde aantal luisteruren per week zal liggen tussen de 1,0 en 2,9 uur. Dit is een ruim interval. Eigenlijk kun je in de praktijk niet zoveel met een dergelijke conclusie. De oorzaak hiervan is de zeer geringe steekproefomvang van 20 elementen. Voor een nauwkeuriger schatting is een veel grotere steekproef nodig.

8.4 Hoe groot moet uw steekproef zijn?

Tijdens het opzetten van een onderzoek komt op een gegeven moment onherroepelijk de vraag naar voren hoe groot de steekproef moet zijn. Dat is een belangrijke beslissing. Immers, als u de steekproef groter neemt dan echt noodzakelijk is, dan

verkwist u veel tijd en geldt. En trek u een te kleine steekproef, dan zullen uw schattingen minder precies zijn dan u had gehoopt.

Het vaststellen van de steekproefomvang is lastig. Er is geen simpele regel die zegt hoe groot een steekproef moet zijn. Al eerder hebben we aangegeven dat er een verband bestaat tussen de omvang van de steekproef en de precisie van de schatting: hoe groter de steekproef, des te preciezer de schatting. Daarom kunt u de vraag naar de steekproefomvang eigenlijk pas beantwoorden als duidelijk is welke precisie u wilt hebben. Dat is dan ook de procedure die we meestal volgen. Eerst stellen we de vereiste precisie vast. Vervolgens kunnen we dan uitrekenen welke steekproefomvang hiervoor nodig is.

Uitgaande van een enkelvoudige aselechte steekproef (met gelijke kansen en zonder teruglegging) zullen we hieronder de procedures beschrijven voor het vaststellen van de steekproefomvang. Dat doen we eerst voor het schatten van percentages, en daarna voor het schatten van het gemiddelde van een variabele.

8.4.1 Steekproefomvang voor het schatten van een percentage

Uitgangspunt bij deze berekeningen is dat de onderzoeker aangeeft hoe groot de onzekerheidsmarge M in de schatting maximaal mag zijn. Deze marge is gelijk aan het verschil tussen de schatting en de bovengrens of ondergrens van het betrouwbaarheidsinterval. We kunnen de onzekerheidsmarge dus opvatten als de maximaal toegestane afwijking van de schatting.

De marge van het 95%-betrouwbaarheidsinterval is gelijk aan de standaardfout vermenigvuldigd met 1,96. Als de marge een bepaalde waarde niet mag overschrijden, dan betekent dat de standaardfout een bepaalde waarde niet mag overschrijden.

De steekproefomvang voor het schatten van een percentage

Als M_{max} de waarde van de onzekerheidsmarge is die niet mag worden overschreden, dan betekent dit dat moet gelden:

$$1,96 \times S(p) \leq M_{max}$$

Invullen van de formule voor de standaardfout van een percentage, en vervolgens uitwerken van deze ongelijkheid, leidt tot de formule

$$n \geq \frac{1}{\frac{N-1}{N} \times \left(\frac{M_{max}}{1,96}\right)^2 \times \frac{1}{P(100-P)} + \frac{1}{N}}$$

Als de omvang N van de doelpopulatie behoorlijk groot is, dan kunnen we bovenstaande formule vereenvoudigen tot

$$n \geq \left(\frac{1,96}{M_{max}}\right)^2 \times P \times (100 - P)$$

Beide formules bevatten het populatiepercentage P . De waarde daarvan is onbekend. Het was immers het doel van het onderzoek die waarde te schatten. Om nu toch de formule te kunnen uitrekenen, kunt u een globale schatting voor P invullen. Dat kan bijvoorbeeld een waarde uit een vorig onderzoek zijn. Als u helemaal geen enkel idee hebt van de waarde van P , vul dan de waarde $P = 50$ in. Dit is de waarde die leidt tot de grootste steekproefomvang. Als voor de waarde $P = 50$ de maximaal toegestane marge niet wordt overschreden, dan zal dat zeker ook niet gebeuren voor andere waarden van P .

Tabel 8.1 geeft voor een reeks waarden van het populatiepercentage en een aantal waarden van de maximaal toegestane onzekerheidsmarge de daarvoor benodigde steekproefomvang. Wilt u een zeer grote precisie (de maximale marge mag bijvoorbeeld niet meer dan 1% zijn), dan betaalt u daarvoor een prijs in de zin van een grote steekproefomvang. Zo is voor het schatten van percentages in de buurt van de 50% binnen een marge van 1% een steekproef nodig van meer dan 9.600 personen. Dit is een wel heel erg grote steekproef. Het zal de kosten van het onderzoek aanzienlijk doen stijgen.

Tabel 8.1. Benodigde steekproefomvang voor het schatten van een percentage

Populatiepercentage	Maximale onzekerheidsmarge				
	1	2	3	4	5
5	1825	457	203	115	73
10	3458	865	385	217	139
15	4899	1225	545	307	196
20	6147	1537	683	385	246
25	7204	1801	801	451	289
30	8068	2017	897	505	323
35	8740	2185	972	547	350
40	9220	2305	1025	577	369
45	9508	2377	1057	595	381
50	9605	2402	1068	601	385
55	9508	2377	1057	595	381
60	9220	2305	1025	577	369
65	8740	2185	972	547	350
70	8068	2017	897	505	323
75	7204	1801	801	451	289
80	6147	1537	683	385	246
85	4899	1225	545	307	196
90	3458	865	385	217	139
95	1825	457	203	115	73

Marktonderzoekbureaus werken vaak met een steekproefomvang in de buurt van de 1.000 personen. Een blik in de kolom voor een maximale marge van 3% leert dat die haalbaar is met zo'n steekproefomvang. Bij een steekproefomvang van 1.068 wordt voor alle populatiepercentages de marge van 3% gehaald. Kortom, met een steekproef van 1.068 personen zal de schatting nooit meer dan 3% van het werkelijke percentage afliggen. Als in de steekproef bijvoorbeeld een schatting voor het percentage luisteraars van 48% wordt gevonden, dan zal het werkelijke percentage

luisteraars in de populatie (met een betrouwbaarheid van 95%) dus liggen tussen 45% en 51%.

8.4.2 Steekproefomvang voor het schatten van een gemiddelde

Uitgangspunt bij deze berekeningen is ook hier weer dat de onderzoeker aangeeft hoe groot de onzekerheidsmarge M in de schatting maximaal mag zijn. Deze marge is gelijk aan het verschil tussen de schatting en de bovengrens of ondergrens van het betrouwbaarheidsinterval. De marge kan dus opgevat als de maximaal toegestane afwijking van de schatting.

De onzekerheidsmarge van het 95%-betrouwbaarheidsinterval is gelijk aan de standaardfout vermenigvuldigd met 1,96. Als de marge een bepaalde waarde niet mag overschrijden, dan betekent dat de standaardfout een bepaalde waarde niet mag overschrijden.

De steekproefomvang voor het schatten van een gemiddelde

Als M_{max} de waarde van de onzekerheidsmarge is die niet mag worden overschreden, dan betekent dit dat moet gelden:

$$1,96 \times S(\bar{y}) \leq M_{max}.$$

Invullen van de formule voor de standaardfout van een gemiddelde, en vervolgens uitwerken van deze ongelijkheid, leidt tot de formule

$$n \geq \frac{1}{\left(\frac{M_{max}}{1,96 \times S}\right)^2 + \frac{1}{N}}$$

Hierin is S de wortel uit de aangepaste populatievariantie S^2 (zie paragraaf 2.4).

Als de omvang N van de doelpopulatie behoorlijk groot is, dan kunnen we bovenstaande formule vereenvoudigen tot

$$n \geq \left(\frac{1,96 \times S}{M_{max}}\right)^2$$

Merk op dat bovenstaande formules de wortel uit de populatievariantie S^2 bevatten. Die waarde is in de praktijk onbekend. Dat maakt het lastig iets met deze formule te doen. Soms is er een indicatie van de waarde van deze grootheid beschikbaar, bijvoorbeeld uit een vorig onderzoek. Dan kunt u deze waarde invullen.

Is er helemaal niets bekend over de waarde van S , dan kunt u nog een andere weg bewandelen. Daarvoor moet u weten welke waarden de variabele ongeveer kan aannemen, en dus ook hoe groot de lengte L van het interval van die mogelijke waarden is. We kunnen nu de volgende situaties onderscheiden:

- De waarden van de variabele hebben een symmetrische, 1-toppige verdeling (de vorm van een berg). Dan zal L ongeveer gelijk zijn aan $6S$, en kan voor S dus de waarde $L / 6$ worden ingevuld.
- De waarden zijn *gelijkmatig* (homogeen) verdeeld over een interval ter lengte L . Elke waarde in het interval komt dus ongeveer even vaak voor. Dan zal S ongeveer gelijk zijn aan $0,3 \times L$.
- De variabele is ongeveer *exponentieel* verdeeld (met veel kleine waarden en weinig grote waarden) over een interval van bekende lengte L . Dan zal S ongeveer gelijk zijn aan $0,4 \times L$.
- De variabele is verdeeld over een interval van bekende lengte L . Over de vorm van de verdeling is *niets bekend*. In het slechtste geval (in termen van variantie) ligt de helft van de waarden bij de ondergrens van het interval, en de andere helft bij de bovengrens. Dan zal S ongeveer gelijk zijn aan $0,5 \times L$.

9. Het probleem van de non-respons

9.1 Non-respons in peilingen

In de voorgaande hoofdstukken hebben we beschreven hoe u een peiling moet opzetten en uitvoeren. Als u zich aan deze leidraad houdt, krijgt u een betrouwbare peiling, waarmee u nauwkeurige uitspraken kunt doen over de populatie die u onderzoekt. Helaas is de praktijk vaak weerbarstiger dan de theorie die we hier tot nu toe hebben beschreven. Bij het uitvoeren van een peiling loopt u altijd tegen praktische problemen aan. Een van de belangrijkste problemen is het optreden van non-respons. Gezien de omvang van dit verschijnsel en de ernstige gevolgen die het kan hebben voor de validiteit van de uitkomsten, gaan we in dit hoofdstuk dieper in op het probleem van de non-respons. We leggen uit wat de effecten van non-respons kunnen zijn op de uitkomsten van uw peiling, en ook hoe u die effecten kunt verminderen.

We spreken van *non-respons* als je in een peiling de gewenste informatie niet krijgt van een persoon, terwijl die persoon toch tot de doelpopulatie van het onderzoek behoort en in de steekproef is getrokken (en dus de gegevens had moeten verstrekken).

We hebben het in dit hoofdstuk over *unit non-respons*. Daarbij krijgt u van een persoon in de steekproef geen enkele informatie. Het vragenformulier blijft dus helemaal leeg. Er is ook nog *item non-respons*. Daarbij blijven alleen enkele vragen onbeantwoord. Het gaat dan vaak over wat gevoelig liggende vragen, bijvoorbeeld over inkomen, zwart geld, seksueel gedrag, gebruik van drugs en crimineel verleden. We zullen het hier niet over item non-respons hebben.

Een eerste, voor de hand liggend, effect van non-respons is dat de gerealiseerde steekproef minder groot is dan u had gepland. Als u gegevens wilt verzamelen met een steekproef van 1.000 personen, en de helft doet niet mee, dan houdt u uiteindelijk maar gegevens van 500 personen over. In principe hoeft dat niet tot onjuiste conclusies te leiden. Het betekent dat uw steekproef kleiner is, en dus is de onzekerheidsmarge, en dus ook het betrouwbaarheidsinterval, groter. U kunt ook al van te voren rekenen houden met non-respons. Als u gegevens van 1.000 personen wilt hebben, en u verwacht dat de helft niet meedoet aan de peiling, trek dan in eerste instantie een (bruto) steekproef van 2.000 personen.

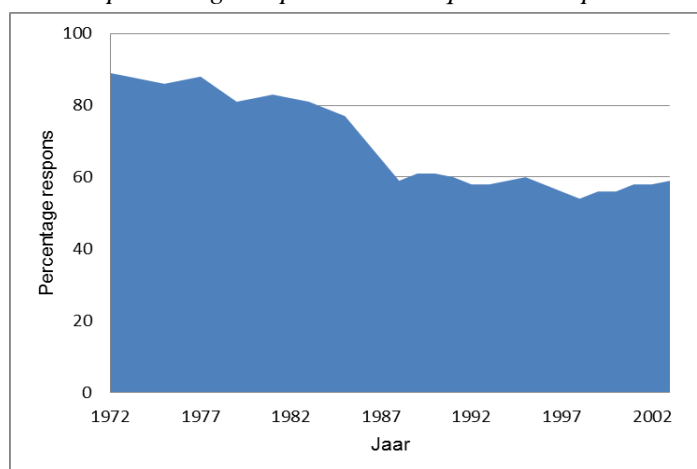
Helaas heeft non-respons meestal ook nog een ander, veel ernstiger, effect. Non-respons kan namelijk *selectief* zijn. Dit verschijnsel doet zich voor als, ten gevolge van non-respons, bepaalde groepen onder- of oververtegenwoordigd zijn in de peiling. Als een ondervertegenwoordigde groep zich duidelijk anders gedraagt met betrekking tot de te onderzoeken variabelen dan de oververtegenwoordigde groep, dan leidt dit tot een *vertekening* in de uitkomsten. Anders gezegd: een schatting valt systematisch te hoog of te laag uit.

Keer op keer blijkt dat non-respons selectief is. Dat was bijvoorbeeld het geval bij een aantal onderzoeken van het Centraal Bureau voor de Statistiek (CBS):

- Bij de Enquête Slachtoffers Misdrijven weigerden bange mensen vaker om mee te doen aan het onderzoek. Daarmee onderschat je dus het percentage bange mensen.
- Aan woningbehoefteonderzoeken deden vooral mensen mee die niet tevreden waren met hun huidige woning. Mensen die tevreden waren met hun huis, hadden minder belangstelling voor zo'n onderzoek.
- Bij het Onderzoek Verplaatsingsgedrag bleek dat mobielere mensen ondervertegenwoordigd waren. Het was lastig om contact met ze te leggen. Ze waren immers minder vaak thuis.
- In verkiezingsonderzoek zijn de mensen die gaan stemmen altijd oververtegenwoordigd.

Non-respons in peilingen neemt eerder toe dan af. Figuur 9.1 toont het verloop van het percentage respons door de jaren heen bij de Enquête Beroepsbevolking (EBB) van het CBS. Bij ander onderzoek zijn dezelfde patronen te zien. In de jaren 70 van de vorige eeuw was de respons nog hoog (bijna 90%), maar in de loop van de tijd is de respons afgekald. Nu mogen we blij zijn met 60% respons. Daarvoor is meestal dan ook nog heel wat inspanning nodig.

Figuur 9.1. Het percentage respons in de Enquête Beroepsbevolking (EBB)



De omvang en het effect van non-respons hangt van heel wat factoren af. In eerste belangrijke factor is het onderwerp van de peiling. Als de benaderde personen belangstelling hebben voor het onderwerp van de peiling, zullen ze vaker geneigd zijn om mee te doen. Bij een saaie, oninteressante peiling is de kans op succes veel kleiner.

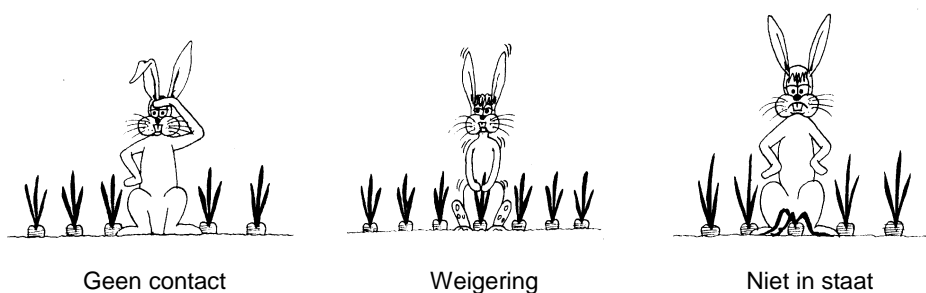
Bij peilingen onder huishoudens is het vaak zo dat één persoon in het huishouden de vragen beantwoordt. Als het er niet zoveel toe doet wie dat is, dan maakt dit de kans op respons een stuk hoger.

Ook de periode waarin u de peiling doet, speelt een rol. U kunt maar beter de vakantieperiodes in de zomer en rondom de kerst mijden. Dan zijn veel mensen niet thuis. En als ze wel thuis zijn, hebben ze het misschien wel te druk met andere zaken. Het verzenden van schriftelijke vragenlijsten in de kerstperiode is niet zo handig. De brief kan dan makkelijker zoekraken tussen alle andere post.

De inzet van enquêteurs heeft heel veel invloed op het percentage respons. Bij mondeling en telefonische peilingen liggen de responspercentages veel hoger dan bij schriftelijke peilingen of peilingen via het internet.

Non-respons kan verschillende oorzaken hebben. Het is goed de non-respons op basis van deze oorzaken in groepen te verdelen. Uit onderzoek van non-respondenten is gebleken dat de diverse groepen nogal kunnen verschillen. Elk type non-respons kan aanleiding geven tot een ander soort vertekening. Een goede indeling van de non-respons is dus onontbeerlijk. Dit geldt niet alleen voor de analyse van de non-respons, maar ook voor een goede verantwoording van het veldwerk is een duidelijke classificatie belangrijk. De belangrijkste drie oorzaken van non-respons zijn 'geen contact', 'weigering' en 'niet in staat'.

Figuur 9.2. Oorzaken van non-respons



De allereerste stap in het verkrijgen respons is het contact leggen met de personen die in de steekproef zijn getrokken. Dat kan om allerlei redenen misgaan. We spreken dan van non-respons door *geen contact*. Als een enquêteur bij iemand aan de deur komt, dan kan deze niet thuis zijn. Dat kan voor korte tijd zijn (een boodschap), voor langere tijd (overwinteren in Spanje), of zelfs permanent (verhuisd naar een onbekend adres, een portier of huismeester weigert de toegang, er loopt een gemene waakhond in de voortuin, of de persoon is overleden). Ook bij een telefonische enquête kan iemand niet thuis zijn. Maar de contactpoging kan ook falen omdat iemand de telefoon niet opneemt of in gesprek is. En bij een schriftelijke peiling kan non-respons ontstaan als het adres onjuist is of de bewoners de post meteen weggooiden. Hetzelfde kan gebeuren bij online-enquêtes als het e-mailadres niet juist is of als de e-mail niet door een spam-filter komt.

Dit soort non-respons kunnen we proberen te verminderen door meer *contactpogingen* te doen. Bij het CBS is het niet ongebruikelijk dat bij een mondelinge enquête een enquêteur zes contactpogingen doet voordat het desbetreffende geval als non-respons wordt afgeboekt.

Is het gelukt om contact te maken met een persoon, dan kunnen we vaststellen of deze behoort tot de doelpopulatie. Is dat niet het geval, dan zijn we klaar. De persoon hoor dan niet thuis in de steekproef en er hoeft daarom geen vragenlijst te worden ingevuld. We kunnen deze persoon negeren als een geval van *overdekking*. Behoort de persoon wel tot de doelpopulatie, dan moeten we deze overhalen om mee te werken aan het onderzoek. Lukt dat niet dan is er sprake van non-respons als gevolg van *weigering*.

Bij weigering is het zinvol om onderscheid te maken tussen tijdelijke weigeraars en permanente weigeraars. Bij tijdelijke weigeraars komt het tijdstip van het gesprek ongelegen. De benaderde persoon heeft op dat moment geen zin of geen tijd (bezig met het klaarmaken van het eten, de baby is ziek, er is een voetbalwedstrijd op TV), maar er is een goede kans dat na het maken van een nieuwe afspraak het gesprek alsnog tot stand komt. Bij permanente weigeraars zal het nooit tot beantwoording van de vragen komen. Permanente weigering kan zich bijvoorbeeld voordoen als de persoon het onderwerp niet aan staat, of als hij vindt dat het onderzoek teveel inbreuk doet op zijn privacy.

Een derde oorzaak van non-respons betreft personen die wel mee willen doen met het onderzoek maar het niet kunnen. Dit kan bijvoorbeeld komen door ziekte, dronkenschap, dementie, doofheid, blindheid of een geestelijke handicap. Ook een taalbarrière tussen de enquêteur en de respondent kan de reden zijn.

9.2 Analyse van de non-respons

We moeten altijd op onze hoede zijn als zich non-respons voordoet in een onderzoek. We kunnen de in hoofdstuk 8 beschreven schattingstechnieken dan niet zonder meer toepassen. We moeten eerst nagaan of de non-respons van dien aard is dat die tot een vertekening kan leiden. En als dat het geval is, dan moeten we een methode vinden om voor deze vertekening te corrigeren.

Het effect van non-respons

Om de effecten van non-respons te kunnen onderzoeken, hanteren we vaak een model waarin we aan ieder object in de populatie een bepaalde (onbekende) kans op respons toe te kennen. Personen die vaak meedoen aan peilingen hebben een grote responskans, en personen die bijna nooit meedoen, hebben een kleine responskans. De responskansen van alle objecten in de populatie noteren we met

$$p_1, p_2, \dots, p_N.$$

Uit deze populatie trekken we een enkelvoudige aselechte steekproef van omvang n . Dan zal niet iedereen in de steekproef meedoen aan de peiling, want er treedt non-respons. We krijgen daarom geen n waarnemingen y_1, y_2, \dots, y_n beschikbaar, maar minder. De omvang van de respons geven we aan met m (waarbij m kleiner dan n) en de beschikbaar gekomen waarden met y_1, y_2, \dots, y_m .

Stel dat we het gemiddelde van een doelvariabele Y willen schatten. Dan ligt het voor de hand (toepassing van het analogieprincipe) om het gemiddelde

$$\bar{y}_R = \frac{1}{m} \sum_{i=1}^m y_i = \frac{y_1 + y_2 + \dots + y_m}{m}$$

van de waarden voor de m responderende objecten als schatter hiervoor te gebruiken. Helaas is dit geen zuivere schatter. Deze schatter heeft een *systematische vertekening*, en die is gelijk aan

$$\frac{R_{Y,p} \times S_Y \times S_p}{p_{gem}}$$

Hier in $R_{Y,p}$ de correlatiecoëfficiënt. Dat is een maat voor de samenhang tussen de doelvariabele en de responskansen. De correlatiecoëfficiënt is alleen 0 als er geen samenhang is. Dus een samenhang draagt bij aan de vertekening.

Verder is S_p de wortel uit de variantie van de responskansen. Naarmate de responskansen meer variëren (er zijn grote kansen en kleine kansen), zal de vertekening groter zijn. Als alle responskansen gelijk zijn, dan is er geen vertekening.

De grootte p_{gem} is het gemiddelde van alle responskansen. Deze grootte kunnen we schatten met de fractie respons in de peiling. Een hoge fractie respons betekent een grote gemiddelde responskans en dus een kleine vertekening. Bij een lage responsfractie is de vertekening groter.

Een diepgaande behandeling van de effecten van en correctie voor non-respons kunt u vinden in Bethlehem, Cobben & Schouten (2011).

Hoe kunt u in de praktijk nagaan of de non-respons selectief is? Met de beschikbare informatie over de doelvariabele (de waarnemingen bij de respondenten) komt u niet verder. U kent immers de waarden voor de non-respondenten niet, zodat u ze niet kunt vergelijken met die van de respondenten. Toch is er een uitweg en dat is gebruik maken van *hulpvariabelen*. Dat zijn variabelen die u hebt gemeten in uw peiling en waarover u nog extra informatie hebt. Denk aan een variabele als geslacht. U kunt van elke respondent het geslacht registreren. Als u dan ziet dat de verhouding man/vrouw in de respons anders is dan in de populatie, dan weet u dat er iets mis is. Als het percentage mannen in de peiling bijvoorbeeld 60% is, dan zitten er dus teveel mannen en te weinig vrouwen in de peiling. Kennelijk responderen mannen dan beter dan vrouwen. Er is dan sprake van een verband tussen responsgedrag en geslacht. Dat leidt tot een selectieve respons.

U moet dus op zoek naar hulpvariabelen die een verband hebben met het responsgedrag. Treft u zulke variabelen aan, dan is de respons selectief en loopt u het risico dat u verkeerde conclusies trekt uit uw peiling.

Mocht de respons selectief zijn, dan kunt u de schattingsprocedures in hoofdstuk 8 niet meer gebruiken. U moet namelijk een correctie uitvoeren voor de selectieve

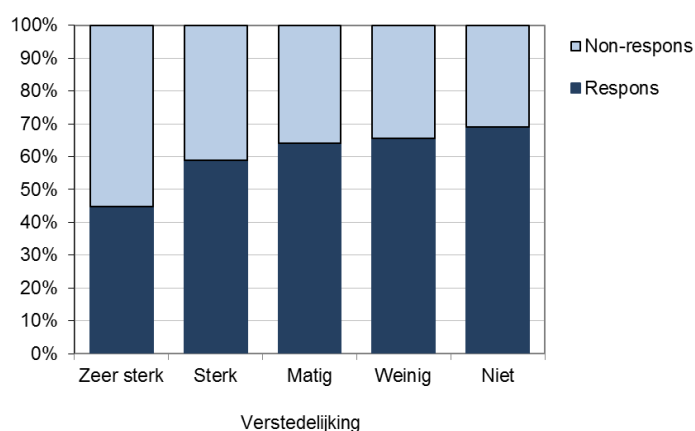
respons. Ook daarvoor kun je de hulpvariabelen gebruiken. Het gaat dan vooral om de hulpvariabelen die een verband hebben met het responsgedrag.

Waar haalt u die hulpvariabelen vandaan? Het gaat om variabelen die u meet in uw peiling en waarvoor u ook de verdeling in de populatie (of in de volledige steekproef) beschikbaar hebt. Hier zijn wat mogelijke bronnen van hulpvariabelen:

- Het steekproefkader. Soms bevat het steekproefkader allerlei variabelen. Een voorbeeld is het bevolkingsregister. Daarin zitten variabelen als geslacht, leeftijd (af te leiden uit de geboortedatum), burgerlijke staat en land van geboorte.
- Het statistisch bureau. Het CBS kan de verdeling van heel wat variabelen leveren. Dat is natuurlijk alleen zinvol voor variabelen die over dezelfde doelpopulatie gaan.
- Observaties van enquêteurs. U kunt hierbij bijvoorbeeld denken aan variabele als het type woning en de ouderdom van de woning.

Figuur 9.3 bevat een voorbeeld van een grafiek waarmee u kunt vaststellen of er verband is tussen responsgedrag en een hulpvariabele. Gegevens zijn afkomstig uit een peiling van het CBS. Het gaat om het Permanent Onderzoek naar de Leefsituatie (POLS) uit 1998. De hulpvariabele is de mate van verstedelijking. Die geeft aan hoe verstedelijkt de plaats is waar de respondent woont.

Figuur 9.3. Verband tussen responspercentage en mate van verstedelijking



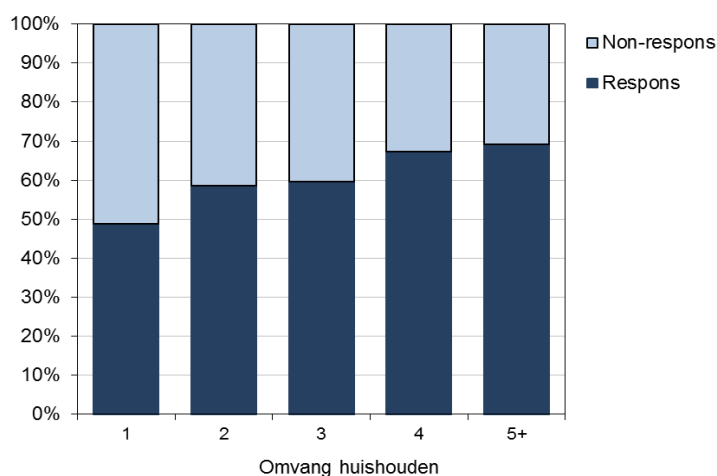
Duidelijk is te zien in de grafiek dat de respons erg laag is in de grote steden (zeer sterk verstedelijkt) en erg hoog op het platteland (niet verstedelijkt). Een dergelijk patroon kom je bijna wereldwijd tegen. Het is lastig om een hoge respons te halen in grote steden, Dat komt vooral omdat het moeilijk is om contact te leggen met de personen die in de steekproef zijn getrokken.

Figuur 9.4 toont nog een andere hulpvariabele. Het gaat om de omvang van het huishouden waarvan de respondent deel uitmaakt.

Ook hier is een duidelijk patroon te zien. De respons neemt toe met de omvang van het huishouden. De belangrijkste oorzaak is dat het vooral bij alleenstaanden moeilijk is om contact te leggen. Kennelijk zijn die vaak niet thuis. Verder is er wat

vaker non-respons bij kleinere huishoudens omdat de personen niet in staat zijn om mee te doen. Het gaat hier vooral om oudere alleenstaanden en echtparen. En ook zijn er wat meer weigeringen bij kleine huishoudens.

Figuur 9.4. Verband tussen responspercentage en omvang van het huishouden



Uit de analyse hierboven blijkt dat mensen in de grote steden en mensen in kleine huishoudens vaak ondervertegenwoordigd zijn in een peiling. Als u iets onderzoekt wat hiermee verband houdt, dan kunt u verwachten dat uw schattingen en vertekening hebben.

Er zijn veel meer hulpvariabelen die een verband hebben met responsgedrag. Zie hiervoor bijvoorbeeld Bethlehem, Cobben & Schouten (2011).

9.3 Correctie voor non-respons

Als de analyse van de non-respons voldoende aanwijzingen oplevert voor een mogelijke vertekening van schattingen, dan is het niet verantwoord is om zonder verdere correcties over te gaan tot publicatie van de uitkomsten. Een veel toegepaste methode om de uitkomsten te corrigeren is het uitvoeren van een *weegprocedure*. Daarbij kent u aan elke waargenomen persoon een *correctiegewicht* toe. In de schattingsprocedures neemt u vervolgens deze correctiegewichten mee.

De effectiviteit van een weegprocedure staat of valt met de beschikbaarheid van geschikte hulpvariabelen. Het gaat om hulpvariabelen die aan twee voorwaarden moeten voldoen:

- Ze moeten een sterke samenhang vertonen met de doelvariabelen van het onderzoek. Is dat niet het geval, dan zal een weging de schattingen voor de doelvariabele niet corrigeren.
- Ze moeten een samenhang vertonen met het responsgedrag. Is dat niet het geval, dan treedt er geen correctie op.

U gebruikt geschikte hulpvariabelen voor het berekenen van *correctiegewichten*. Die gewichten bepaalt u zodanig dat de gewogen verdeling van de variabele in de respons exact gelijk wordt aan die in de populatie. U maakt de respons zo

representatief met betrekking tot de hulpvariabele. Dit bereikt u door ondervertegenwoordigde groepen een hoger gewicht te geven en oververtegenwoordigde groepen een lager gewicht.

Als het mogelijk is om de steekproef tegelijk representatief te maken met betrekking tot een aantal hulpvariabelen, en die hulpvariabelen hangen allemaal sterk samen met de doelvariabelen van het onderzoek, dan zal de (gewogen) steekproef ook (bij benadering) representatief zijn met betrekking tot de doelvariabelen. Daardoor zullen schattingen voor de doelvariabele gebaseerd op de gewogen steekproef beter zijn dan schattingen die zijn gebaseerd op de ongewogen steekproef.

We gebruiken een eenvoudig voorbeeld om wegen te illustreren. De bevolking van het (denkbeeldige) land Samplonië bestaat uit (slechts) 1.000 zielen. Van het Samplonische CBS weten we dat er 511 mannen en 489 vrouwen zijn. We zouden dus het geslacht als hulpvariabele kunnen gebruiken. We doen een peiling. De uiteindelijke respons blijkt te bestaan uit 100 personen, waarvan 48 mannen en 52 vrouwen. Zie ook tabel 9.1

Tabel 9.1. Het wegen van de respons met de hulpvariabele geslacht

Steekproef			Populatie			Correctiegewicht	
	Aantal	Perc		Aantal	Perc		
Man	48	48,0%	Man	511	51,1%	Man	1,065
Vrouw	52	52,0%	Vrouw	489	48,9%	Vrouw	0,940
Totaal	100	100,0%	Totaal	1000	100,0%		

Uit de tabel blijkt dat de verhouding man/vrouw in de steekproef anders is dan in de populatie: de steekproef bestaat voor 48% uit mannen en in de populatie is dat 51%. We kunnen nu de steekproef representatief maken met betrekking tot de variabele geslacht door de mannen een correctiegewicht te geven dat gelijk is aan

$$\frac{\text{Percentage mannen in de populatie}}{\text{Percentage mannen in de steekproef}} = \frac{51,1}{48,0} = 1,065 .$$

Op dezelfde wijze krijgen de vrouwen een correctiegewicht

$$\frac{\text{Percentage vrouwen in de populatie}}{\text{Percentage vrouwen in de steekproef}} = \frac{48,9}{52,0} = 0,940 .$$

Dat de mannen een correctiegewicht groter dan 1 krijgen, is niet verwonderlijk. Ze zijn immers ondervertegenwoordigd in de respons. In feite telt nu elke man in de respons mee voor 1,065 man. Vrouwen zijn oververtegenwoordigd en krijgen een correctiegewicht kleiner dan 1. Elke vrouw in de respons telt mee voor 0,940 vrouw. Zouden we nu op grond van de steekproef het percentage mannen willen schatten dan vinden we

$$100 \times (48 \times 1,065) / 100 = 51,1$$

en dat is precies de fractie mannen in de populatie. Evenzo komt de fractie vrouwen precies goed uit. De gewogen steekproef is dus representatief met betrekking tot de variabele geslacht.

Zijn er meer hulpvariabelen, dat wordt de zaak ingewikkelder. Ook dan kunnen we een indeling in groepen maken op basis van de hulpvariabelen. We moeten dan de hulpvariabelen met elkaar kruisen. Hebben we één hulpvariabele, dan zijn er evenveel groepen als de hulpvariabele categorieën heeft. Bij meer variabelen is het aantal groepen het product van de aantallen categorieën per variabele. Stel we hebben bij het voorbeeld ook nog de beschikking over de hulpvariabele leeftijd in drie categorieën (jong, middelbaar en oud). Dan is er bij wegen naar de hulpvariabelen leeftijd en geslacht een groep voor elke combinatie van leeftijd en geslacht. Dat zijn hier dus $2 \times 3 = 6$ groepen. Kennen we nu de verdeling van de populatie over de aldus gevormde groepen, dan kunnen we voor elke groep een correctiegewicht bepalen.

Tabel 9.2. Wegen met twee hulpvariabelen

Steekproef			Populatie			Correctiegewicht		
	Man	Vrouw		Man	Vrouw		Man	Vrouw
Jong	23	15	Jong	226	209	Jong	0,983	1,393
Middel	16	17	Middel	152	144	Middel	0,950	0,847
Oud	13	16	Oud	133	136	Oud	1,023	0,850

Tabel 9.2 laat zien hoe dit werkt de hand van een steekproef van omvang 100. De gewichten zijn op dezelfde manier bepaald als in tabel 9.1. Zo krijgen oude vrouwen en gewicht van 0,850. Dit is het resultaat van deling van 13,6 door 16,0. We hebben nu bereikt dat de steekproef representatief is met betrekking tot zowel leeftijd als geslacht. Sterker nog, de steekproef is ook representatief voor geslacht binnen elke leeftijdscategorie en, omgekeerd, voor leeftijd binnen elk geslacht.

Naarmate u meer relevante hulpvariabelen gebruikt in uw weging, zal de vertekening kleiner zijn. Houdt daarbij in de gaten dat de groepen die u krijgt door het kruisen van hulpvariabelen, aan de volgende twee voorwaarden moeten voldoen:

- De groepen moeten *homogeen* zijn met betrekking tot de *doelvariabelen* van je peiling. Dat betekent dat alle personen binnen een groep voor wat betreft de doelvariabele zoveel mogelijk op elkaar moeten lijken. Binnen de groepen moet er zo weinig mogelijk variatie in de waarden van de doelvariabele zitten.
- De groepen moet *homogeen* zijn met betrekking tot het *responsgedrag*. Dat betekent dat alle personen binnen een groep ongeveer dezelfde kans moeten hebben om aan het onderzoek mee te doen (te responderen).

In de praktijk is het lang niet altijd eenvoudig om geschikte hulpvariabelen te vinden. Dat betekent dat u het moet doen met de hulpvariabelen die u hebt. Bedenk dan dat correctie voor non-respons minder effectief zal zijn. Mogelijk is de vertekening in de uitkomsten wel minder, maar niet helemaal verdwenen.

Voorbeeld 9.1. Correctie voor non-respons in een luisteronderzoek

In de gemeente Rijnwoude is ooit een luisteronderzoek uitgevoerd. Een van de onderzoeksvragen was hoeveel inwoners van de gemeente naar de lokale omroep luisteren. De doelpopulatie bestond uit alle inwoners van de gemeenten van 12 jaar en ouder.

De oorspronkelijke steekproef bestond uit 499 personen. Die kwamen uit een steekproef van adressen. Uiteindelijk deden 209 personen mee aan het de peiling. Daarmee was het percentage respons gelijk aan $100 \times 209 / 499 = 41,9\%$. Dat is niet hoog. Daarom alleen al moesten de onderzoekers rekening houden met een mogelijke vertekening in de uitkomsten.

Van de bevolking van 12 jaar en ouder was de verdeling over geslacht en leeftijdsgroep bekend. De gemeente had deze bevolkingsstatistieken beschikbaar. De vragen naar geslacht en leeftijd waren opgenomen in de vragenlijst. Daardoor kon de percentuele verdeling over de groepen voor de respons worden vergeleken met die van de populatie:

Respons (n=209)			Populatie (N=19950)			Correctiegewicht		
	Man	Vrouw		Man	Vrouw		Man	Vrouw
Jong	9,5%	18,3%	Jong	12,5%	13,0%	Jong	1,136	0,613
Middel	17,7%	28,6%	Middel	26,3%	27,7%	Middel	1,284	0,835
Oud	13,7%	12,3%	Oud	8,9%	11,6%	Oud	0,561	0,817

Er zijn behoorlijke verschillen tussen respons en populatie. Vrouwen zijn in alle leeftijdsgroepen oververtegenwoordigd in het onderzoek. Mannen van middelbare leeftijd zijn duidelijk ondervertegenwoordigd in het onderzoek. Dat zou te maken kunnen hebben met het feit dat ze overdag werken, waardoor het moeilijker is om contact te maken.

Merk op dat de percentages voor de respons zijn berekend met inachtneming van de ongelijke trekkingskansen in het onderzoek. Schattingen voor de doelvariabelen bevatten dus in feite twee gewichten: het gewicht dat corrigeert voor de ongelijke trekkingskansen en het gewicht dat corrigeert voor selectieve non-respons.

Het effect van al die correctieprocedures laten we zien aan de hand van de vraag of men wel eens naar de lokale omroep luistert.

Luistert u wel eens naar de lokale omroep?	
In de ongecorrigeerde steekproef:	55,0 %
Na correctie voor ongelijke trekkingskansen:	59,1 %
No correctie voor non-respons:	57,8 %

In de ongecorrigeerde respons is het percentage luisteraars gelijk aan 55,0%. Correctie leidt uiteindelijk tot een hoger percentage luisteraars van 57,8%. Dit zal een betere schatting zijn.

10. Analyse van de gegevens

10.1 Vormen van analyse

Na afronding van het veldwerk heeft u een grote hoeveelheid ingevulde formulieren. Heeft u gewerkt met papieren vragenlijsten, dan moet u de gegevens op die formulieren invoeren in de computer, waarna u ze kunt gaan analyseren. Die analyse kan verschillende vormen aannemen.

Een eerste stap is het uitvoeren van een exploratieve analyse. *Exploratieve analyse* richt zich op het verkennen van een (vaak grote) verzameling gegevens en het in kaart brengen van de karakteristieke eigenschappen daarvan. Het is een verkennende analyse, die we uitvoeren in een situatie waarin we nog geen duidelijk beeld hebben van wat er aan de hand is. Daarom zullen we behoefte hebben aan technieken die ons helpen zoeken naar patronen en structuren in de gegevens. Het is vooral belangrijk dat deze technieken in staat zijn onverwachte aspecten in de gegevens naar voren te brengen.

De eerste fase van de exploratieve analyse is het controleren van de gegevens. De analyse kan mogelijke problemen in de gegevens aan het licht brengen. Zo zou een sterk afwijkende waarde van een variabele (een *uitschieter*) kunnen duiden op een fout. En ook ongebruikelijk combinaties van waarden van twee variabelen (iemand van 12 jaar die heeft gestemd bij de laatste verkiezingen) moet leiden tot nadere inspectie van het desbetreffende formulier. Zo kan de exploratieve analyse een rol spelen in de controle en correctie van de gegevens (zie ook hoofdstuk 7).

De tweede fase van de exploratieve analyse voert u uit na controle en correctie. Het is een verkennende inhoudelijke analyse. Daarin gaat u op zoek naar interessante patronen en structuren in de gegevens. De exploratieve analyse biedt u hiervoor een scala aan technieken om de karakteristieke eigenschappen van vooral grote hoeveelheden gegevens op compacte wijze in een beperkt aantal kengetallen, tabellen of grafieken vast te leggen. Het zou mooi zijn als u zo onverwachte aspecten in de gegevens zou ontdekken. Daar neemt immers uw kennis over het desbetreffende onderwerp toe.

Met de exploratieve analyse onderzoekt u alleen maar de verzamelde gegevens. U kijkt wat er in de gegevens aan de hand is en uw conclusies hebben alleen maar betrekking op de verzamelde gegevens zelf. U wilt echter meestal meer met een peiling. U heeft een steekproef uit een populatie getrokken en u wilt op basis van die steekproef conclusies kunnen trekken over de populatie als geheel. U wilt dus generaliseren van de steekproef naar de populatie. Dan komen we op het terrein van de *inductieve analyse*. Daarin nemen uw uitspraken meestal de vorm aan van beweringen over karakteristieke kenmerken van de populatie als geheel. Het zouden beweringen kunnen zijn over het gemiddelde inkomen van een bevolkingsgroep, het percentage stemgerechtigden dat denkt te gaan stemmen, of de totale mestproductie

in Nederland. Het kan dan gaan om het maken van schattingen van deze grootheden, of om het toetsen van beweringen over de waarde van deze grootheden. Omdat we uitspraken doen op grond van waarnemingen in een steekproef en niet in de gehele populatie, hebben ze een element van onzekerheid. De inductieve analyse biedt echter middelen om die onzekerheid te kwantificeren.

Er zitten wat haken en ogen aan het maken van schattingen van populatie-grootheden. Dat komt omdat uw gegevens meestal niet zo mooi zijn als u hoopt dat ze zijn. Er zijn allerlei zaken waarmee u rekening moet houden:

- Uw steekproef is niet met gelijke, maar met ongelijke kansen getrokken. Dan zijn de gegevens niet representatief. Om goede cijfers te krijgen, moet u voor die ongelijke kansen corrigeren in uw schattingen.
- U heeft fouten in de antwoorden (of ontbrekende antwoorden) gecorrigeerd door ze te vervangen door synthetische waarden (imputatie). Als personen met ontbrekende antwoorden verschillen van personen die wel antwoorden, dan kan dit leiden tot onjuiste schattingen. En als u, bijvoorbeeld, imputatie van het gemiddelde hebt toegepast, dan leveren uw berekeningen voor onzekerheidsmarges te kleine waarden op, waardoor u onterecht denkt dat u schattingen heel precies zijn.
- U heeft geen garantie dat de door de respondenten gegeven antwoorden ook inderdaad de juiste antwoorden zijn. Respondenten kunnen een sociaalwenselijk antwoord geven. Of u doet teveel een beroep op de herinnering van de respondenten waardoor ze zaken vergeten te melden. En ook kunnen respondenten vluchten naar het antwoord 'weet niet'.
- U krijgt te maken met non-respons. Als de non-respondenten afwijken van de respondenten (en dat gebeurt vaak), dan loopt u het gevaar de verkeerde conclusies uit uw peiling te trekken. U zult moeten wegen om hiervoor te corrigeren. In uw schattingen zult u dan correctiegewichten mee moeten nemen.

Als u iets bijzonders ontdekt, zult u het ook willen kunnen verklaren. Dat betekent dat u op zoek gaat naar relaties tussen variabelen. Daarvoor zijn allerlei geavanceerde analysetechnieken beschikbaar zoals regressieanalyse en factoranalyse. Dit soort technieken vallen buiten het bestek van deze publicatie.

In de rest van dit hoofdstuk richten we ons vooral op het gebruik van technieken voor exploratieve analyse. We onderscheiden technieken die de verdeling van één variabele analyseren en technieken die de samenhang tussen variabelen onderzoeken. Verder moet u zich realiseren dat u voor de analyse van kwantitatieve variabelen andere technieken moet gebruiken dan voor de analyse van kwalitatieve variabelen. Tabel 10.1 geeft een overzicht van de technieken die aan bod komen. Dit is geen uitputtend overzicht. Er zijn nog veel meer technieken, maar we beperken ons hier tot de belangrijkste.

Tabel 10.1. Technieken voor exploratieve analyse

Variabelen	Analyse van de verdeling	Analyse van de samenhang
Kwantitatief	Spreidingsdiagram Snorrendoos Histogram Numeriek overzicht	Puntenwolk Correlatiecoëfficiënt
Kwalitatief	Staafdiagram Cirkeldiagram Frequentieverdeling	Samengesteld staafdiagram Stapeldiagram Kruistabel
Gemengd		Analyse van de verdeling van de kwantitatieve variabele voor elke categorie van de kwalitatieve variabele

Er zijn grafische en numerieke technieken voor exploratieve analyse. Het altijd nuttig om met grafische technieken te beginnen. Het Chinese spreekwoord dat één plaatje meer zegt dan 1000 woorden, gaat hier zeker op. Plaatjes kunnen een grote hoeveelheid informatie op overzichtelijke wijze weergeven en zo allerlei inzichten verschaffen. Blijkt uit de grafieken dat er sprake is van duidelijke, simpele patronen, dan kunt daarna numerieke technieken gebruiken om de die structuren in enkele kengetallen vast te leggen.

Veel van de hier besproken technieken zijn terug te vinden in statistische programmatuur zoals *SPSS*, *SAS* en *Stata*. Dat zijn overwegend grote en dure pakketten die veel meer kunnen dan nodig is voor exploratieve analyse. Hier is gekozen voor een simpele aanpak waarbij we gebruik maken van Excel (om de gegevens in op te slaan) en *R* (een gratis te downloaden open source pakket voor statistische analyse).

De verschillende analysetechnieken illustreren we aan de hand van voorbeelden. Daarvoor maken we gebruik van een bestand met gegevens over 341 werkende inwoners van het denkbeeldige land *Samplonië*. De volgende variabelen zijn gemeten:

- *Gemeente*: kwalitatieve variabele met 7 categorieën: Akkerwinde, Grasmalen, Nieuwekans, Lommerdal, Smeulde, Stapelrade en Vuilpanne.
- *Provincie*: kwalitatieve variabele met 2 categorieën: Agrië en Indusië.
- *Geslacht*: kwalitatieve variabele met 2 categorieën: Man en Vrouw.
- *Leeftijd*: kwantitatieve variabele met waarden tussen de 20 en 64.
- *Werkzaam*: indicatorvariabele voor het hebben van werk, 1=Werk, 0=Werkloos.
- *Inkomen*: kwantitatieve variabele met waarden tussen de 101 en 4497.
- *Leeftijdsklasse*: kwalitatieve variabele met 3 categorieën: Jong, Middelbaar en Oud, afgeleid van Leeftijd.

Om de verzamelde gegevens met de computer te kunnen analyseren, moeten we ze eerst in de computer invoeren. We hebben hiervoor het spreadsheetprogramma Excel gebruikt. Figuur 10.1 toont een deel van de spreadsheet met de gegevens over *Samplonië*.

Figuur 10.1. Spreadsheet met de ingevoerde gegevens

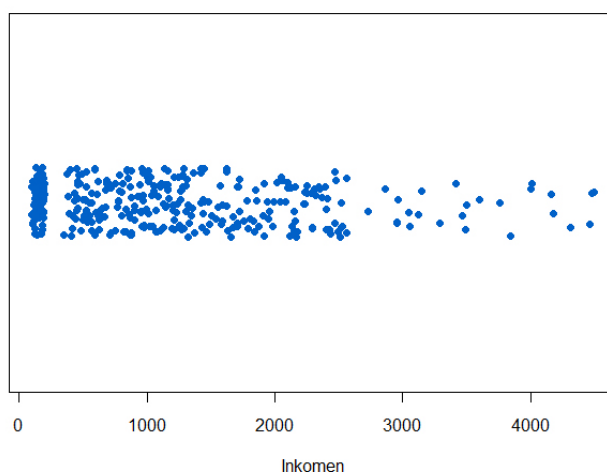
	A	B	C	D	E	F	G
1	Gemeente	Provincie	Geslacht	Leeftijd	Werkzaam	Inkomen	LeeftKlas
2	Nieuwekans	Agrië	Vrouw	33	Werk	158	Jong
3	Akkerwinde	Agrië	Man	32	Werk	1525	Jong
4	Vuilpanne	Indusië	Vrouw	63	Werk	1442	Oud
5	Stapelrade	Indusië	Man	29	Werk	1126	Jong
6	Akkerwinde	Agrië	Vrouw	33	Werk	195	Jong
7	Vuilpanne	Indusië	Man	39	Werk	1559	Middelbaar
8	Lommerdal	Indusië	Man	54	Werk	3760	Oud
9	Nieuwekans	Agrië	Man	40	Werk	630	Middelbaar
10	Vuilpanne	Indusië	Man	53	Werk	2103	Oud

Het statistische pakket R kunt u downloaden van de website www.r-project.org. Nadat u R hebt geïnstalleerd op uw computer, slaat u eerst uw spreadsheet op als csv-bestand. Vervolgens kunt u in R dit csv-bestand weer inlezen.

10.2 Analyse van de verdeling van een kwantitatieve variabele

We beginnen met drie grafische technieken. Dat zijn het spreidingsdiagram, de snorrendoos en het histogram. Daarna bespreken we nog het numerieke overzicht van de verdeling.

Figuur 10.2. Een spreidingsdiagram



Het spreidingsdiagram brengt de verdeling in zijn meest pure vorm in beeld. Op een horizontale as wordt een schaalverdeling aangebracht en op die schaalverdeling worden de waarden als individuele punten afgezet. Figuur 10.2 toont een voorbeeld van een spreidingsdiagram. Hierin is de variabele *Inkomen* afgezet voor de werkende bevolking van *Samplonië*. Er is in verticale zin een beetje ruis toegevoegd aan de gegevens. Daardoor worden punten uit elkaar getrokken en zichtbaar gemaakt die elkaar anders zouden overlappen.

Waarop moet u letten bij een spreidingsdiagram? Daarvoor zijn moeilijk algemene regels te geven, omdat u altijd bedacht moet zijn op het onverwachte zaken. Toch zijn hier wat aspecten:

- *Uitschieters*. Zijn er waarnemingen die buiten de boot vallen, die zich zeer afwijkend van de rest gedragen? Zulke waarnemingen manifesteren zich als losse, geïsoleerde punten. Dergelijke vreemde eenden in de bijt moeten altijd

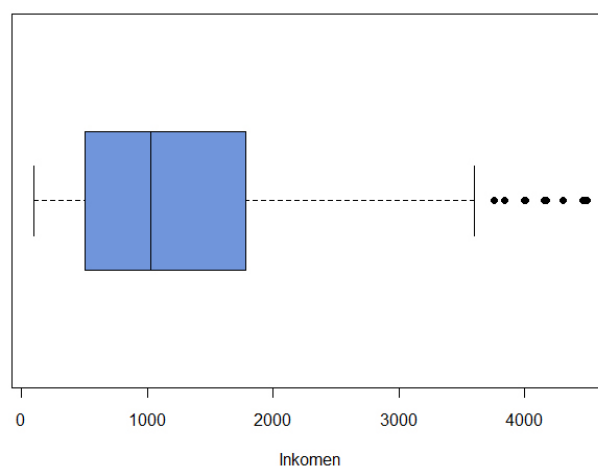
even goed worden bekeken. Misschien horen die waarnemingen niet thuis bij de gegevens. Of misschien zijn er fouten gemaakt bij het invoeren van de gegevens in de computer. Het kan echter ook zijn dat u uiteindelijk besluiten dat de waarde correct is, zodat u die dus gewoon moet meenemen in de analyse. Voorzichtigheid is in ieder geval geboden.

- *Groepering*. Liggen de waarnemingen verspreid over het hele gebied, of vallen er verschillende groepjes te onderscheiden? Als u een aantal groepjes ziet, kan dit erop duiden dat de waarnemingen afkomstig zijn uit verschillende populaties die door elkaar zijn gemengd. Het zou dan beter kunnen zijn om die groepjes apart te analyseren.
- *Concentratie*. Is er sprake van een bepaald gebied waar de dichtheid van de punten hoog is? Misschien concentreren de waarnemingen zich misschien wel rondom een locatie. Als dat het geval is, dan is het belangrijk deze locatie nader te karakteriseren. Hiervoor kunt u dan weer andere technieken gebruiken.

In figuur 10.2 zijn geen uitschieters te zien. Er lijkt wel een aparte groep te zijn met heel lage inkomens. Verder onderzoek zal moeten uitwijzen wat dit precies voor personen zijn. De inkomens lijken zich niet erg om één specifiek waarde te concentreren. Er is sprake van een scheve verdeling met veel lage en weinig hoge inkomens.

Een tweede grafische techniek voor het in beeld brengen van de verdeling van een kwantitatieve variabele is de *snorrendoos*. Snorrendoos is de Nederlandse vertaling van *box-and-whisker plot*. Een snorrendoos karakteriseert de verdeling door een doos waar aan beide zijden snorharen uitsteken. Figuur 10.3 bevat de snorrendoos van de inkomens van de werkende in *Samplonië*.

Figuur 10.3. Een snorrendoos



De snorrendoos bestaat uit een rechthoekige doos. Deze doos geeft het gebied aan waarbinnen precies de middelste helft (50%) van de waarnemingen ligt. De verticale streep in de doos duidt de mediaan aan, de middelste waarde. Van de linker en rechter zijkant van de doos lopen lijnen ('snorharen') naar de zogenaamde 'naburige waarden'. Dit zijn de waarnemingen die nog net binnen een afstand van 1,5 maal de lengte van de doos van de linker- resp. rechterkant van de doos liggen. Alle

waarnemingen die verder weg liggen, worden apart getekend. Dat zijn de uitschieters.

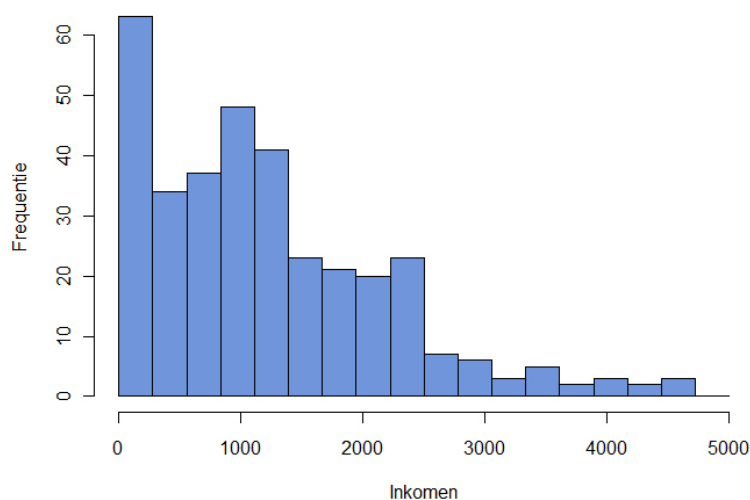
De snorrendoos kan u helpen bij het opsporen van uitschieters. Dat zijn in principe de punten die in de grafiek los zijn getekend. Maar voorzichtigheid is geboden. Aangezien de verdeling in figuur 10.3 erg scheef is, waarbij de rechterstaart nogal lang doorloopt, is hier eigenlijk geen sprake van uitschieters. De scheefheid van de verdeling kunnen we goed aflezen uit de snorrendoos zelf: de streep staat niet in het midden van de doos, en ook de linker snorhaar is veel korter dan de rechter snorhaar.

De traditionele vorm om de verdeling van een kwantitatieve variabele weer te geven is het *histogram*. Om een histogram te kunnen tekenen moet u het waardebereik van de variabele eerst in een aantal klassen verdelen, en voor elke klasse moet het aantal waarden daarin tellen. Vervolgens verdeelt u de horizontale as in intervallen die overeenkomen met deze klassen. Boven elk interval wordt een kolom getekend waarvan de oppervlakte evenredig is met het aantal waarnemingen in die klasse. De kolommen moeten zo worden getekend dat ze elkaar raken. Er mag geen tussenruimte zijn.

Een punt van overweging is de keuze van het aantal klassen. Met te weinig klassen kan de grafiek te grof zijn en valt er amper nog iets te zien aan de vorm van de verdeling. Met teveel klassen komen er teveel details in te grafiek naar voren die het algemene beeld van de verdeling verstoren. Een veel gebruikte vuistregel zegt dat het aantal klassen ongeveer gelijk moet zijn aan de wortel uit het aantal waarnemingen, waarbij u een minimum van 5 en een maximum van 20 in acht moet nemen.

Aan de hand van het histogram kunt u vaststellen of de verdeling symmetrisch en eentoppig is. Als dat het geval is, dan kunt u met een numeriek overzicht locatie en spreiding van de verdeling van de variabele in een paar simpele kengetallen vastleggen.

Figuur 10.4. Een histogram



Figuur 10.4 bevat een voorbeeld van een histogram. De verdeling van het inkomen van de werkende personen in *Samplonië* is in beeld gebracht. Er is sprake van een

zeer asymmetrische verdeling. Dit doet zich vaak voor bij het meten van omvangen of hoeveelheden. Er zijn veel mensen die weinig verdienen en weinig mensen die veel verdienen.

Mocht er sprake zijn van meer dan één top in de verdeling, dan zou dit kunnen wijzen op een mengsel van meerdere verdelingen. De waarnemingen zijn dan afkomstig uit verschillende groepen die elk een eigen, andere verdeling hebben. Het kan dan zinvol zijn die groepen te identificeren en apart te analyseren.

In figuur 10.4 lijkt er sprake te zijn van enkele bergen en dalen. Nader onderzoek naar een opsplitsing in groepen is dus wellicht de moeite waard. Het zal duidelijk zijn dat er bij het optreden van meertoppigheid het lastig is om de verdeling te karakteriseren door één locatie.

Als de verdeling van de waarnemingen er netjes (symmetrisch en eentoppig) uitziet, dan kunt u al die waarden redelijk samenvatten in de vorm van een numeriek overzicht. Een dergelijk overzicht zou de volgende kengetallen kunnen bevatten.

- *Minimum*. Dat is de kleinste waarde die voorkomt.
- *Maximum*. Dat is de grootste waarde die voorkomt.
- *Gemiddelde*. Dat is de centrale locatie van de verdeling waar omheen alle waarden gespreid liggen.
- *Standaardafwijking*. Dat is een maat voor de spreiding. Naarmate de standaardafwijking groter is, zit er ook meer variatie in de waarden.
- *Vuistregelinterval*. Dat is een interval waarin ongeveer 95% van de waarden ligt (mits de verdeling eentoppig en symmetrisch is). De ondergrens van dit interval is het gemiddelde min twee keer de standaardafwijking, en de bovengrens is het gemiddelde plus twee keer de standaardafwijking.

De verdeling van de inkomens in *Samplonië* is scheef en meertoppig, en leent zich daarom niet zo goed om samenvatten te vatten. Als we ons echter beperken tot alleen de 58 werkende mannen in de provincie Agrië, dan is de verdeling wel netjes. De numerieke samenvatting hiervan staat in tabel 10.2.

Tabel 10.2. Numeriek overzicht

Variabele	Inkomen
Aantal waarnemingen	58
Minimum:	353
Maximum:	841
Gemiddelde:	551,2
Standaardafwijking:	119,3
Vuistregelinterval:	(312,6 ; 789,8)

De inkomens van de werkende mannen in *Agrië* liggen kennelijk tussen 353 en 841, waarbij ze geconcentreerd zijn rondom de gemiddelde waarde 551,2. Een standaardafwijking van 11,3 leidt tot een vuistregelinterval van 312,6 tot 789,8.

10.3 Analyse van de verdeling van een kwalitatieve variabele

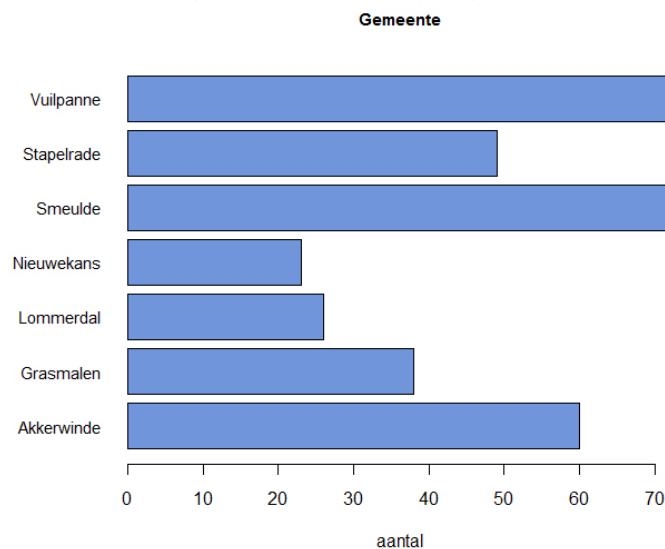
Voor het onderzoeken van de verdeling van een kwalitatieve variabelen zijn slechts weinig technieken beschikbaar. Dat wordt voor een belangrijk deel veroorzaakt door

het feit dat we met een kwalitatieve variabele niet echt kunnen rekenen. Zo'n variabele maakt slechts een verdeling in categorieën. Het enige wat u met die indeling kunt doen, is de aantallen waarnemingen per categorie met elkaar vergelijken. Dat geldt ook voor de grafische technieken. We zullen hier slechts twee technieken bespreken: het staafdiagram en het cirkeldiagram.

De eerste techniek is het *staafdiagram*. Hierbij worden de categorieën weergegeven als staven, waarbij de lengtes van de staven overeenkomen met de aantallen waarnemingen per categorie. Dus hoe langer de staaf des te meer waarnemingen in de categorie. Om niet de indruk van een kwantitatieve verdeling te wekken (zoals bij een histogram), is het beter om de staven horizontaal en los van elkaar te tekenen.

Figuur 10.5 bevat een staafdiagram van de aantallen werkenden in de zeven gemeenten van *Samplonië*. Voor elke gemeente is er een staaf, en de lengte van elke staaf is evenredig met het aantal werkenden in die gemeente. Uit de figuur blijkt dat twee gemeenten duidelijk meer werkende personen hebben dan de andere gemeenten: *Vuilpanne* en *Smeulde*. Ook de twee kleine gemeenten *Lommerdal* en *Nieuwekans* vallen op.

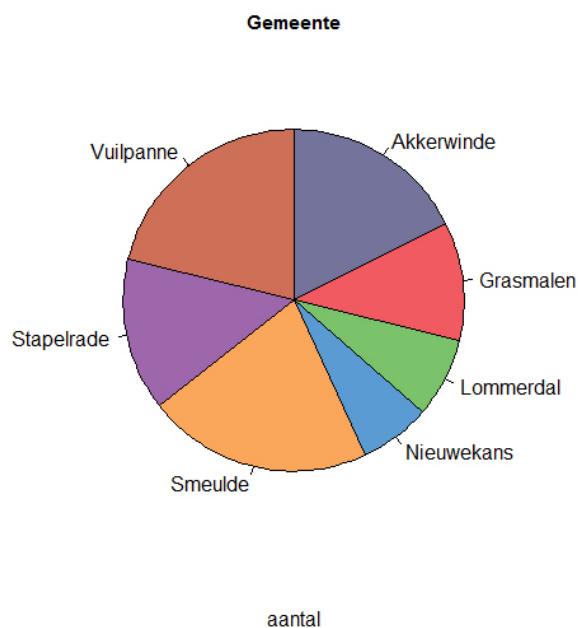
Figuur 10.5. Een staafdiagram



Vooraf in de media is een andere grafische weergave van de verdeling populair, en dat is het *cirkeldiagram*. Hierbij verdelen we een cirkel in evenveel parten (sectoren) als er categorieën zijn. De oppervlakte van deze parten nemen we evenredig aan de aantallen waarnemingen in de categorieën. Figuur 10.6 bevat het cirkeldiagram van de inwoneraantallen in de gemeenten van *Samplonië*. Hij bevat dus dezelfde informatie als figuur 10.5, maar dan op andere wijze weergegeven.

Cirkeldiagrammen hebben soms het nadeel dat ze wat minder makkelijk zijn te interpreteren. Als er relatief veel parten zijn, en die parten ontlopen elkaar niet al te veel in omvang, dan is de onderlinge vergelijking lastig. In dat opzicht zijn staafdiagrammen wat makkelijker te interpreteren.

Figuur 10.6. Een cirkeldiagram



De numerieke wijze van weergeven van de verdeling van een kwalitatieve variabele is de frequentieverdeling. Dit is een tabel waarin voor elke categorie zowel het aantal als het percentage waarnemingen staat vermeld. Tabel 10.3 bevat de frequentieverdeling voor de aantallen werkende personen per gemeente in *Samplonië*.

Tabel 10.3. Een frequentieverdeling

Categorie	Aantal	Percentage
Akkerwinde	60	17,6%
Grasmalen	38	11,1%
Lommerdal	26	7,6%
Nieuwekans	23	6,7%
Smeulde	73	21,4%
Stapelrade	49	14,4%
Vuilpanne	72	21,1%
Totaal	341	100,0%

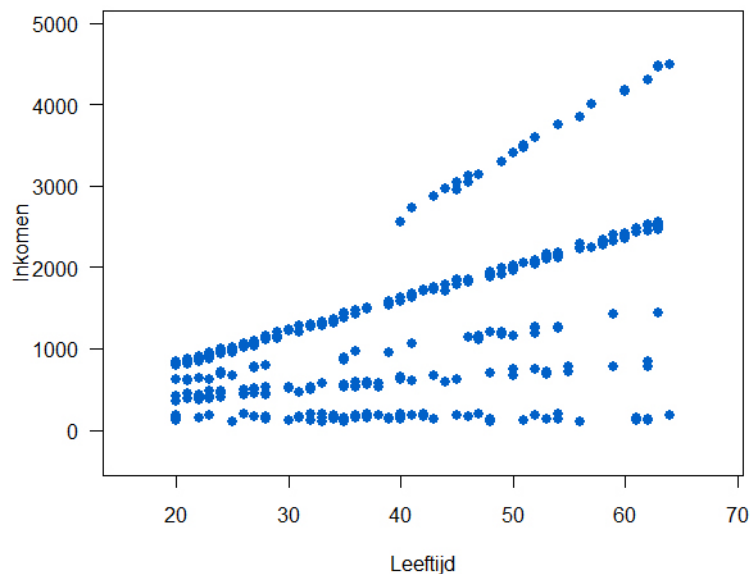
Met een kwalitatieve variabele kan niet echt worden gerekend. Daarom is er niet zoiets als een gemiddelde. Als dan toch in de vorm van een kengetal iets over de verdeling moet worden gezegd, dan komt daarvoor de modus in aanmerking. De *modus* is gedefinieerd als de categorie waarin de meeste waarnemingen zitten. Voor de inwoneraantallen is dat de categorie Smeulde met 73 werkende personen, wat neer komt op 21,4% van het totaal.

10.4 Analyse van de samenhang tussen kwantitatieve variabelen.

Voor de grafische analyse van de samenhang tussen twee kwantitatieve variabelen is de *puntenwolk* de meest voor de hand liggende techniek. Dat is een grafiek waarin we elke persoon als een punt weergeven. De horizontale coördinaat van een punt is gelijk aan de waarde van de ene variabele en de verticale coördinaat is gelijk aan de waarde van de andere variabele.

Zijn er duidelijke patronen te herkennen in de puntenwolk, dan is er sprake van een bepaalde vorm van samenhang. Deze moet u dan uiteraard verder onderzoeken en verklaren. De meest extreme vorm van samenhang is die waarbij alle punten op een rechte lijn liggen. In dat geval kan de waarde van de ene variabele exact worden voorspeld uit de waarde van de andere variabele. Maar ook andere aspecten kunnen worden ontdekt in een puntenwolk. Zo zal het vrij simpel zijn om uitschieters te detecteren. En ook afwijkende groepjes waarnemingen zullen duidelijk opvallen. Heeft de puntenwolk de vorm van een ongestructureerde sneeuwstorm, dan is er geen samenhang.

Figuur 10.7. Een puntenwolk



Figuur 10.7 bevat de puntenwolk van de leeftijd tegen inkomen voor de werkende bevolking van *Samplonië*. Er vallen twee dingen op in deze puntenwolk. In de eerste plaats vallen de punten uit elkaar in verschillende groepen, en in de tweede plaats is er binnen de groepen sprake van een redelijk rechtlijnig verband. In de groep met de hogere inkomens loopt het inkomen op met de leeftijd, en in de groep met de laagste inkomens lijkt het inkomen min of meer onafhankelijk te zijn van de leeftijd.

Zeker als de puntenwolk in groepen uit elkaar valt, is het interessant te onderzoeken wat dat precies voor groepen zijn. Dat kan betekenen dat u er meer variabelen bij gaat betrekken. Het is bijvoorbeeld mogelijk dat grootte van de punten te laten afhangen van de waarde van een derde (kwantitatieve) variabele, of de kleur van de punten te laten bepalen door de categorieën van een derde (kwalitatieve) variabele.

De structuur in een puntenwolk kan vrij ingewikkeld zijn. Het lijkt daar om niet zo waarschijnlijk u dit hele beeld in één getal of enkele getallen kunt vastleggen. In sommige situaties kan dat echter wel. Als het verband tussen beide variabelen rechtlijnig is (de punten liggen ongeveer op een rechte lijn), dan kunt u de samenhang samenvatten met de correlatiecoëfficiënt en de regressielijn.

De *correlatiecoëfficiënt* (of meer voluit de *productmoment correlatiecoëfficiënt*) probeert de sterkte van de samenhang uit te drukken in een getal dat ligt tussen -1 en

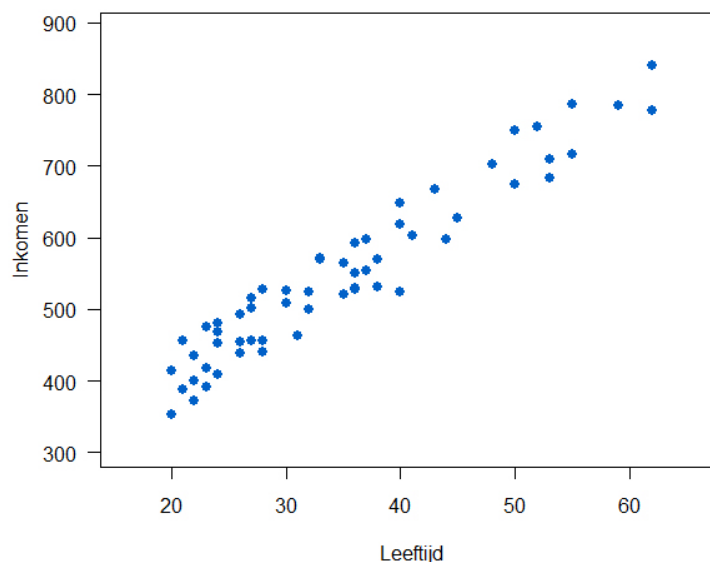
+1. Daarbij geldt wel de beperking tot deze grootte alleen in staat is om rechtlijnige samenhang te meten. Liggen de punten bijvoorbeeld heel mooi op een parabolisch kromme, dan is er wel sprake van een sterke samenhang, maar zal de correlatiecoëfficiënt dat niet oppikken.

De waarde van de correlatiecoëfficiënt kan variëren tussen -1 en +1. Als er totaal geen samenhang is tussen twee variabelen, dan is de waarde van de correlatiecoëfficiënt gelijk aan nul. Liggen de waarden exact op een rechte lijn, dan is de correlatie gelijk aan +1 (stijgende lijn) of -1 (dalende lijn). In die situatie kan de waarde van de ene variabele exact worden voorspeld als de waarde van de andere variabele bekend is.

Het is niet zo zinvol om een correlatiecoëfficiënt uit te rekenen zonder eerst naar de puntenwolk zelf te hebben gekeken. Dat blijkt wel uit figuur 10.7. De waarde van de correlatiecoëfficiënt is niet erg hoog: 0,568. Toch lijkt er in elke groep apart wel een sterke samenhang te zijn. Zou u bijvoorbeeld de correlatie uitrekenen voor alleen de bovenste groep met hoge inkomens, dan zou u een waarde vinden van 0,964, hetgeen inderdaad duidt op een zeer sterke samenhang binnen die groep.

Als er sprake is van een redelijk sterke samenhang, en deze samenhang is bovendien rechtlijnig van karakter, dan kunt u ook nog de aard van de samenhang nader beschrijven. Het gaat dan om het vinden van de formule van de rechte lijn die zo goed mogelijk door de punten heen gaat. Een dergelijke lijn noemen we een *regressielijn*.

Figuur 10.8. Een puntenwolk met een rechtlijnige samenhang



De berekeningen voor de regressielijn vallen buiten het bestek van deze publicatie. Met de programmatuur voor statistische analyse is dit echter meestal simpel te doen. Voor de gegevens in figuur 10.7 is het niet zinvol om een regressielijn te berekenen, aangezien ze niet (bij benadering) op ene rechte lijn liggen. Anders ligt dat als we ons beperken tot de werkende mannen in provincie *Agrië*. De puntenwolk voor het verband tussen leeftijd en inkomen staat in figuur 10.8. Hier is (bij benadering) sprake van een rechtlijnige samenhang. De correlatiecoëfficiënt is dus een goede

maat voor de sterkte van de samenhang. Die correlatie is gelijk aan 0,960. Die waarde ligt vlak bij 1 en dus is er een sterke samenhang. De formule voor de regressielijn is $205,493 + 9,811 \times \text{Leeftijd}$. Dus u kunt het inkomen van iemand goed voorspellen door zijn leeftijd te vermenigvuldigen met 9,811 en daar dan nog eens 205,493 bij op te tellen.

10.5 Analyse van de samenhang tussen kwalitatieve variabelen.

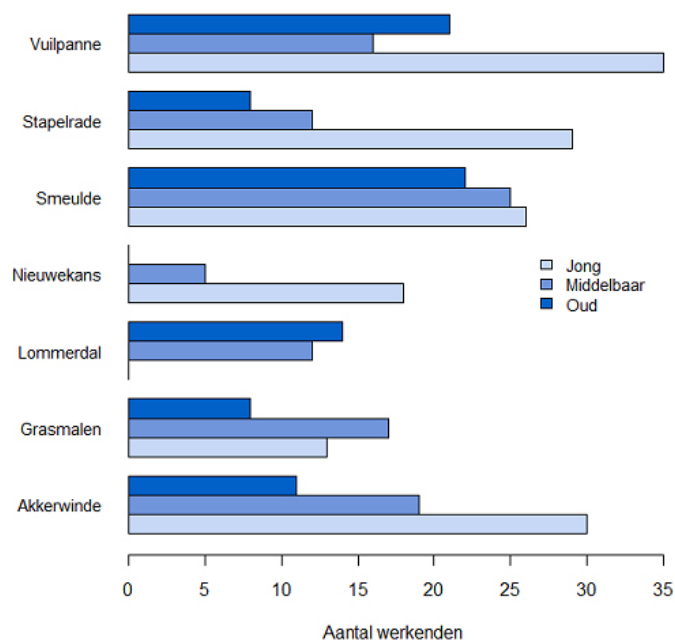
De mogelijkheden voor het onderzoek van de samenhang tussen kwalitatieve variabelen is beperkt. Dit heeft te maken met de al eerder genoemde onmogelijkheid om met de waarden van kwalitatieve variabelen berekeningen uit te voeren. Voor de grafische analyse kunt u gebruik maken van enkele variaties op het staafdiagram. Dat zijn het samengesteld staafdiagram en het stapeldiagram. Verder kunt u gebruik maken van cirkeldiagrammen.

Bij het *samengesteld staafdiagram* tekent u voor elke categorie van de ene kwalitatieve variabele een staafdiagram van de andere variabele. Al die staafdiagrammen tekent u naast elkaar in dezelfde figuur. Figuur 10.9 bevat een voorbeeld van een dergelijke grafiek. Het betreft de leeftijdsverdeling (in klassen) van de werkende bevolking in de gemeenten van Samplonië.

De staven zijn, net zoals bij een simpel staafdiagram, horizontaal getekend. Daarmee wordt de suggestie van een histogram vermeden.

In de grafiek is een staafdiagram van de leeftijdsklassen getekend voor elke gemeente. Indien u dat wenst is, kunt u natuurlijk ook het omgekeerde doen: een staafdiagram van de verdeling over de gemeenten voor elke leeftijdscategorie.

Figuur 10.9. Samengesteld staafdiagram



Is de vorm van het staafdiagrammen anders voor elke categorie, dan is er sprake van samenhang tussen de twee variabelen. Hoe die samenhang dan in elkaar zit, kunt u

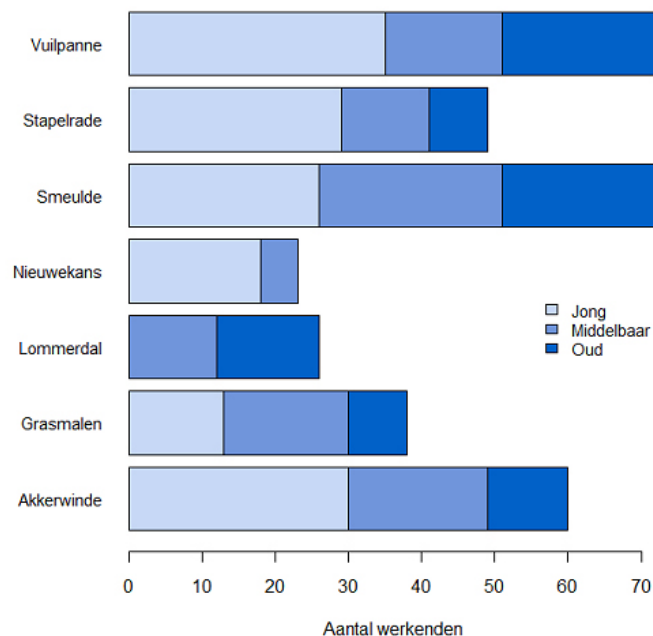
nagaan door nader onderzoek van de staafdiagrammen. Sommige aspecten kunnen wel goed worden afgelezen uit de grafiek en andere aspecten niet. Zo is moeilijk te zien welke gemeente nu de grootste is. Daarvoor zouden we immers alle staven van het staafdiagram voor voor elke gemeente op elkaar moeten stapelen. Ook is moeilijk af te lezen of een leeftijdscategorie relatief gezien over- of ondervertegenwoordigd is in een gemeente. De vraag of het percentage jongeren in *Smeulde* nu groter of kleiner is dan het percentage in *Vuilpanne*, is niet zonder meer te beantwoorden. De absolute omvang van de leeftijdscategorieën zijn wel goed af te lezen. Zo wonen in *Smeulde* de meeste ouderen.

Een andere manier om de staafdiagrammen van de ene variabele voor elk categorie van de andere variabele in beeld te brengen, is het *stapeldiagram*. Hierbij worden de staven van één staafdiagram niet onder elkaar gezet maar achter elkaar geplakt. Figuur 10.10 geeft een voorbeeld van een dergelijk stapeldiagram.

De staven zijn zo getekend dat hun lengtes overeenkomen met de aantallen werkende inwoners in de diverse gemeenten. Op hun beurt zijn de staven verdeeld in segmenten die verdeling over de drie leeftijdsklassen in de juiste verhouding weergeven.

Wat valt er te zien aan een stapeldiagram? Net zoals bij een simpel staafdiagram wordt duidelijk welke categorie van de ene variabele (hier: gemeente) het grootste is en welke het kleinste. Verder kan ook aardig worden ingeschat welke categorie van de andere variabele (hier: leeftijdsklasse) relatief goed of slecht is vertegenwoordigd binnen een categorie van de ene variabele. Zo is duidelijk te zien dat er geen jonge werkende zijn in *Lommerdal* en geen oudere werkenden in *Nieuwekans*. Het vergelijken van de leeftijdsverdeling van twee gemeenten is vrij lastig.

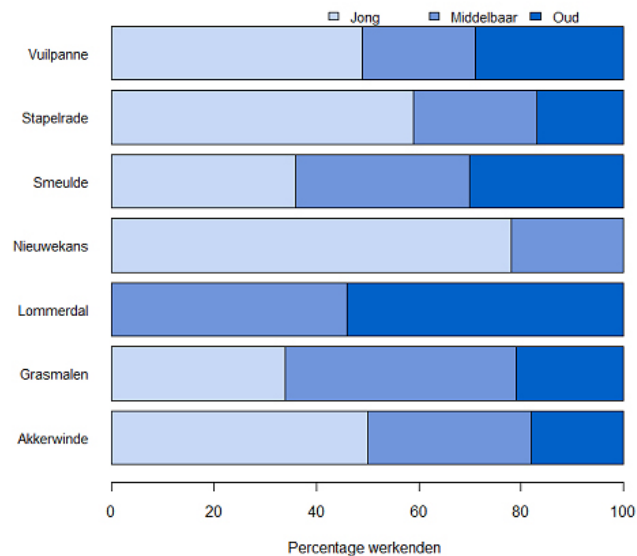
Figuur 10.10. Een stapeldiagram



Een manier om wat meer inzicht te krijgen in dit laatste aspect, is het stapeldiagram waarbij de staven tot 100 zijn gepercenteerd. Dat betekent dat alle staven even lang

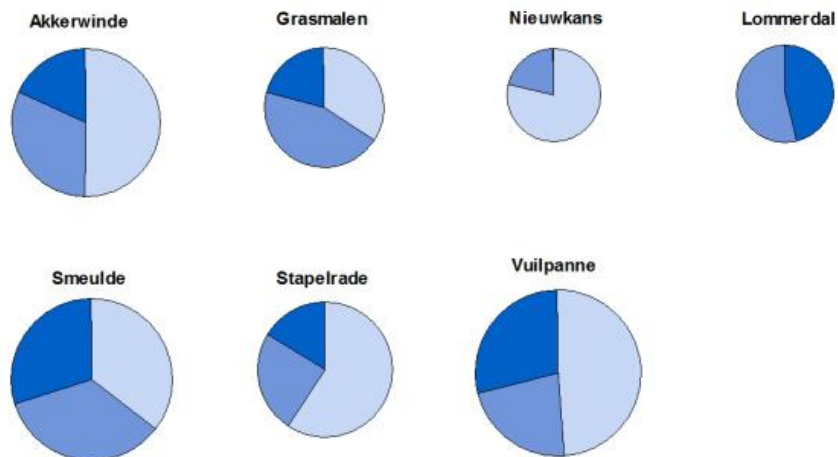
worden (100%), maar binnen de staven worden de onderlinge verhoudingen van de andere variabele nog steeds goed weergegeven. In voorbeeld staat in figuur 10.11.

Figuur 10.11. Een stapeldiagram, met staven die optellen tot 100%



Nu kan ook de relatieve vertegenwoordiging van ouderen beter worden beoordeeld. Duidelijk is bijvoorbeeld te zien dat in *Smeulde* en *Vuilpanne* relatief meer ouderen wonen dan in *Akkerwinde* of *Grasmalen*. Relatief veel jongeren wonen er in *Nieuwekans* en *Stapelrade*, terwijl mensen van middelbare leeftijd goed vertegenwoordigd zijn in *Grasmalen*.

Figuur 10.12. Cirkeldiagrammen



Al eerder hebben we het cirkeldiagram genoemd als alternatief voor het staafdiagram. Ook voor het onderzoek van de samenhang tussen twee kwalitatieve variabelen zou u cirkeldiagrammen kunnen gebruiken. De meest simpele aanpak is om de staven in figuur 10.9 te vervangen door de bijbehorende cirkeldiagrammen. Nog informatiever is het om met de grootte van de cirkels de omvang van de categorieën van de andere variabele aan te geven. Dat is gebeurd in figuur 10.12.

De cirkels zijn zo getekend dat hun oppervlaktes evenredig zijn met de aantallen inwoners in de bijbehorende gemeenten. Uit de grafiek kunt u niet alleen aflezen welke gemeenten klein zijn en welke groot, maar ook hoe de leeftijdsopbouw binnen de gemeenten is. Het is hier wat lastiger om de omvang van bepaalde leeftijdsgroepen voor verschillende gemeenten met elkaar te vergelijken.

Er zijn dus verschillende manieren om de samenhang tussen twee kwalitatieve variabelen te onderzoeken. Er springt niet één techniek uit als de beste. Elke type grafiek geeft bepaalde aspecten van de samenhang weer. Daarom is het waarschijnlijk het beste om al deze grafieken uit te proberen op de gegevens.

U kunt de gezamenlijke verdeling van twee kwalitatieve variabelen numeriek weergeven in een *kruistabel*. Dat is de tweedimensionale versie van de frequentieverdeling die we al in paragraaf 10.3 hebben besproken. Tabel 10.4 toont een voorbeeld van een kruistabel. Hij bevat de verdeling van de variabelen *Gemeente* en *Leeftijdsklasse* voor de werkend bevolking van *Samplonië*.

Tabel 10.4. Een kruistabel

Gemeente	Leeftijdsklasse			
	Jong	Middelbaar	Oud	Totaal
Akkerwinde	30	19	11	60
Grasmalen	13	17	8	38
Lommerdal	0	12	14	26
Nieuwekans	18	5	0	23
Smeulde	26	25	22	73
Stapelrade	29	12	8	49
Vuilpanne	35	16	21	72
Totaal	151	106	84	341

Een relatief kleine tabel als deze is nog wel overzichtelijk, maar voor heel grote tabellen met veel rijen en kolommen kan de interpretatie nog wel eens lastig zijn. Wat nog zou kunnen helpen is het overgaan van aantallen op percentages. Dat kan op verschillende manieren: percentages van het tabeltotaal, rijpercentages en kolompercentages. In tabel 10.5 zijn rijpercentages berekend. Dat betekent dat de percentages in de rijen optellen tot 100%. Zo krijgt u de verdeling van de leeftijdsklassen binnen elke gemeente apart.

Tabel 10.5. Een kruistabel met rijpercentages

Gemeente	Leeftijdsklasse			
	Jong	Middelbaar	Oud	Totaal
Akkerwinde	50,0%	31,7%	18,3%	100%
Grasmalen	34,2%	44,7%	21,1%	100%
Lommerdal	0,0%	46,2%	53,8%	100%
Nieuwekans	78,3%	21,7%	0,0%	100%
Smeulde	35,6%	34,2%	30,1%	100%
Stapelrade	59,2%	24,5%	16,3%	100%
Vuilpanne	48,6%	22,2%	29,2%	100%
Totaal	44,3%	31,1%	24,6%	100%

In de tabel is bijvoorbeeld te zien dat de jongeren oververtegenwoordigd zijn in Nieuwekans (78,3%) en relatief veel ouderen wonen in Lommerdal (53,8%).

Als er geen samenhang is tussen de rij- en kolomvariabele, dan zullen enerzijds de relatieve verdelingen binnen de rijen op elkaar lijken en anderzijds zullen ook de relatieve verdelingen binnen de kolommen op elkaar lijken. Is dat niet het geval dan bestaat er op zijn minst enige samenhang.

Er bestaan numerieke grootheden die trachten de sterkte van de samenhang in de tabel in één getal te vangen. De meest bekende grootheid is de *chi-kwadraat-grootheid*. Als deze een waarde dicht bij 0 heeft, dan is er geen enkele samenhang. Naarmate de waarde van de chi-kwadraat-grootheid toeneemt, is er sprake van meer samenhang. Het probleem bij de chi-kwadraat-grootheid is dat de waarde ervan ook afhangt van het aantal waarnemingen in de tabel en het aantal rijen en kolommen. Daardoor kunt u nooit eenvoudig vaststellen wanneer de samenhang sterk is en wanneer zwak. Een betere grootheid is Cramér's V . Dat is een soort gestandaardiseerde chi-kwadraat-grootheid. De waarde van Cramér's V ligt altijd tussen 0 en 1. Als de waarde gelijk is aan 0, betekent dit totaal geen samenhang, terwijl een waarde van 1 op perfecte samenhang duidt. Soms wordt een vuistregel gehanteerd waarbij men waarden tot 0,3 aanduidt als zwakke samenhang, waarden tussen 0,3 en 0,7 als redelijke samenhang, en waarden boven de 0,7 als sterke samenhang. In het voorbeeld van tabel 10.3 is de waarde van Cramér's V gelijk aan 0,268. Daaruit kunt u concluderen dat er sprake is van slechts een zwakke samenhang tussen de variabelen Gemeente en Leeftijdsklasse.

10.6 De samenhang tussen gemengde variabelen

Voor de analyse van de samenhang tussen een kwantitatieve en een kwalitatieve variabele bestaan geen specifieke technieken. Wel kunt u gebruik maken van al eerder besproken technieken. Hiervoor kiest u een techniek voor het weergeven van de verdeling van een kwantitatieve variabele en die techniek past u vervolgens toe voor elke categorie van de kwalitatieve variabele.

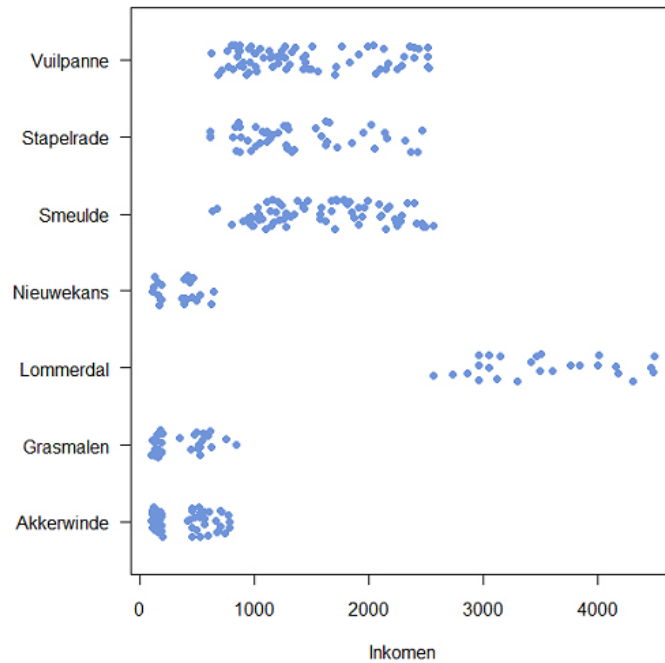
In deze paragraaf bespreken we twee grafische technieken. Ze zijn gebaseerd op het spreidingsdiagram en de snorrendoos. Verder kunnen we ook nog een numeriek overzicht maken.

We beginnen met een techniek gebaseerd op het spreidingsdiagram. Het idee is om in één grafiek spreidingsdiagrammen te maken van de kwantitatieve variabele voor elke categorie van de kwalitatieve variabele. Om te voorkomen dat er teveel punten over elkaar heen vallen, en zo de interpretatie vertroebelen, voegen we een beetje verticale ruis toe.

Figuur 10.13 bevat een voorbeeld van deze aanpak. Hierin wordt het verband onderzocht tussen de kwalitatieve variabele *Gemeente* en de kwantitatieve variabele *Inkomen* (voor de werkenden in *Samplonië*).

Duidelijk is in de grafiek te zien dat in de gemeenten *Akkerwinde*, *Grasmalen* en *Nieuwekans* de inkomens erg laag zijn. Bovendien lijken er in deze gemeenten twee gescheiden groepen te zijn. In gemeente *Lommerdal* zijn de inkomens erg hoog. De overige drie gemeenten nemen een middenpositie in.

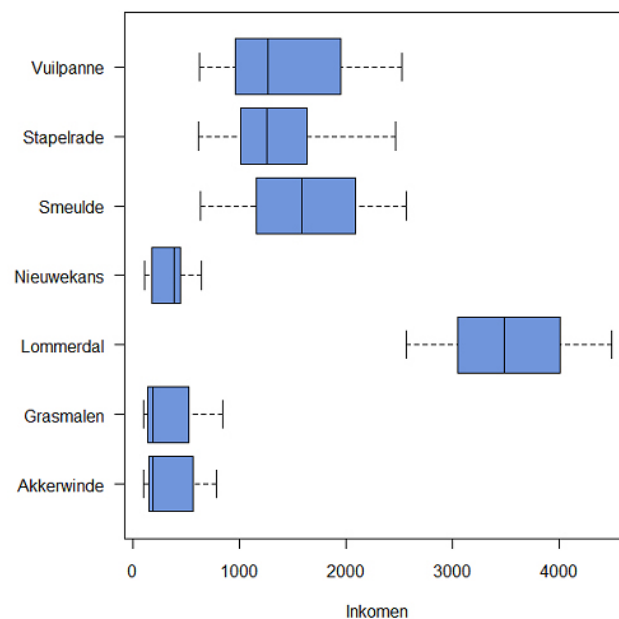
Figuur 10.13. Spreidingsdiagrammen



De tweede grafische techniek die we hier bespreken, is de snorrendoos. Daarin is voor elke categorie van de kwalitatieve variabele een snorrendoos gemaakt van de verdeling van de kwantitatieve variabele. Door al die snorrendozen in één grafiek te tekenen, met dezelfde schaalverdeling, kunnen we de verdelingen onderling vergelijken.

Figuur 10.14 toont deze snorrendozen voor de analyse van de samenhang tussen inkomen en gemeente van de werkende bevolking in Samplonië. In deze grafiek komen de inkomensverschillen duidelijk naar voren. Er zijn kennelijk gemeenten waar de inkomens laag zijn (Akkerwinde, Grasmalen en Nieuwekans) en er is ook een gemeente met hele hoge inkomens (Lommerdal). De inkomensverdeling in de overige gemeenten ligt daar tussen in, en is voor die gemeenten ongeveer hetzelfde.

Figuur 10.14. Snorrendozen



Voor de numerieke analyse van de samenhang tussen een kwalitatieve en een kwantitatieve variabele kunt u dezelfde benadering kiezen als bij de grafische analyse: U gebruikt een numerieke techniek voor een kwantitatieve variabele binnen elke categorie van de kwalitatieve variabele.

Tabel 10.6 bevat het overzicht met daarin de grootheden minimum, maximum, gemiddelde en standaardafwijking. Andere grootheden zijn denkbaar, zoals bijvoorbeeld de mediaan. Ook hier valt al snel de afwijkende inkomensverdeling in *Lommerdal* op. De standaardafwijking van het inkomen in de drie gemeenten *Nieuwekans*, *Grasmalen* *Akkerwinde* is kleiner dan in de overige gemeenten. Kennelijk liggen in deze drie gemeenten de inkomens dicht bij elkaar.

Tabel 10.6. Een numeriek overzicht

Gemeente	Aantal	Minimum	Maximum	Gemiddelde	St. afwijking
Akkerwinde	60	101	787	356	234
Grasmalen	38	102	841	324	219
Lommerdal	26	2564	4497	3534	586
Nieuwekans	23	115	648	344	167
Smeulde	73	635	2563	1607	518
Stapelrade	49	612	2471	1356	505
Vuilpanne	72	625	2524	1440	569
Totaal	341	101	4497	1234	964

We bevelen aan om eerst de vorm van de inkomensverdelingen per gemeente te controleren alvorens u ze samenvat in een tabel. Deze samenvatting is alleen zinvol als de verdelingen er netjes uitzien (symmetrisch, klokvormig en met een piek in het midden). Dat is hier op zijn minst twijfelachtig, aangezien de verdelingen nogal scheef zijn. Een consequentie daarvan is dat de ondergrens van het vuistregelinterval soms te laag is. Er ontstaan negatieve ondergrenzen terwijl we zeker weten dat de waarden altijd positief moeten zijn.

11. Publicatie van de resultaten

De uitkomsten van de analyse van de gegevens die u in uw peiling hebt verzameld, zult u uiteindelijk op een of andere manier willen publiceren. Het ligt voor de hand dit te doen in de vorm van een onderzoeksrapport. In dit hoofdstuk beschrijven we de vorm en inhoud van zo'n rapport.

In het onderzoeksrapport doet u verslag van opzet, uitvoering, analyse en uitkomsten uw peiling. Het rapport moet aan twee belangrijke eisen voldoen. In de eerste plaats moet u de uitkomsten van uw onderzoek in voor de lezers begrijpelijke taal beschrijven. U moet daarbij letten op uw taalgebruik, en technisch jargon zoveel mogelijk vermijden. In de tweede plaats moeten u als onderzoeker in het rapport verantwoording afleggen over de manier waarop u het onderzoek heeft opgezet en uitgevoerd. Dat moet u zo doen dat andere experts op het gebied van peilingen kunnen beoordelen of uw de juiste conclusies heeft getrokken over de doelpopulatie.

U moet het onderzoeksrapport schrijven in een korte, bondige en zakelijke schrijfstijl. Het rapport moet objectief en neutraal zijn, en niet een bepaalde mening opdringen. Het gebruik van spreektaal, en ook de je-stijl of u-stijl, wordt afgeraden. Ook moet u onbekende termen of symbolen vermijden.

U kunt het onderzoeksverslag op allerlei manieren schrijven. Vaak wordt echter een bepaalde structuur geadviseerd. In die structuur zou het verslag uit de volgende onderdelen kunnen bestaan:

- *Samenvatting.* Een korte samenvatting van het doel van de peilingen en de uitkomsten. Deze samenvatting moet zijn geschreven in voor leken begrijpelijke taal.
- *Opzet en uitvoering.* Dit is het methodologische deel van het verslag. Hierin beschrijft u nauwkeurig hoe u uw peiling hebt opgezet en uitgevoerd.
- *Uitkomsten.* In dit deel beschrijft u de uitkomsten van uw analyse. Het zal een mix zijn van tekst, tabellen en grafieken.
- *Conclusies.* Hier trekt u conclusies uit uw onderzoek. U interpreteert de uitkomsten en vertaalt het resultaat naar de praktijk.
- *Literatuur.* Een overzicht van de literatuur die u heeft geraadpleegd. Het kan gaan om zowel methodologische als inhoudelijke publicaties.
- *Bijlagen.* Hierin kunt allerlei technische details over het onderzoek opnemen. Hierbij kunt u bijvoorbeeld denken aan de vragenlijst en tabellen die te groot zijn om in de tekst op te nemen.

11.1 Deel 1: de samenvatting

De *samenvatting* (*executive summary*) geeft een korte beschrijving van het onderzoek in voor de lezers leesbare en begrijpelijk termen. Deze samenvatting valt uiteen in twee onderdelen: de probleemstelling en de conclusies.

De probleembeschrijving geeft een overzicht van het doel van het onderzoek. Dat overzicht zou kunnen beginnen met een verhaal in algemene, inhoudelijk termen. Daarna geeft u concreet en systematisch aan hoe u in uw peiling het antwoord heeft proberen te vinden vragen die voortvloeien uit de probleemstelling.

In de beschrijving van de probleemstelling moet u ook duidelijk maken wie de opdrachtgever is van de peiling, en wie de peiling betaalt.

In het tweede deel van de samenvatting geeft u een overzicht van de belangrijkste conclusies die u hebt getrokken op basis van de uitkomsten van de peiling. Deze conclusies dienen niets meer (maar ook niets minder) te bevatten dan de gevolgtrekkingen die zijn gemaakt, en die direct betrekking hebben op de geformuleerde probleemstelling. Het is belangrijk dat u de conclusies in begrijpelijke taal formuleert. Ook moet u duidelijk aangeven wat de draagwijdte van de conclusies is. U moet voorkomen dat de uitkomsten verkeerd worden geïnterpreteerd. Dat betekent dat u aangeeft op welke doelpopulatie de uitkomsten betrekking hebben, hoe groot de respons was en welke onzekerheidsmarges we minimaal in acht moeten nemen

De samenvatting zal in het algemeen kort zijn en uit niet meer dan een paar pagina's bestaan. Vooral conclusies moet u kort en bondig presenteren, uiteraard in de juiste context. Hier is geen plaats voor de argumentatie die heeft geleid tot de conclusies.

Niet-statistisch onderlegde lezers moeten de samenvatting goed kunnen lezen en begrijpen. Een opdrachtgever zou met de samenvatting in de hand verantwoorde beleidsbeslissingen moeten kunnen nemen. In dit deel van het rapport is geen behoefte aan wiskundige of statistische hoogstandjes.

11.2 Deel 2: de methodologische verantwoording

Het tweede deel van het onderzoeksrapport is de methodologische verantwoording van de peiling. De beschrijving van de opzet en uitvoering van de peiling moet voldoende informatie bevatten om te kunnen vaststellen of de getrokken conclusies ook inderdaad correct zijn. Minimaal moet u de volgende zaken vermelden.

- Een exacte beschrijving van de *doelpopulatie*. U moet duidelijk maken wie wel en wie niet tot de doelpopulatie behoren, en op welke groep mensen de conclusies van het onderzoek dus betrekking hebben.
- De *variabelen* die u heeft gemeten in de peiling. U moet bij kwalitatieve variabelen beschrijven welke categorieën u heeft onderscheiden, en bij kwantitatieve variabelen wat de meeteenheid is.

- Informatie over de *vragenlijst*, zoals aantal vragen en de tijd die het kostte om hem in te vullen. Verder moet u duidelijk maken of het een papieren of een elektronische vragenlijst was. Geef vervolgens of er controles in de vragenlijst waren opgenomen. En beschrijf in het kort hoe de vragenlijst is getest. De vragenlijst zelf neemt u op in de bijlagen.
- Geef een overzicht van de *populatiekenmerken* die u hebt geschat. Leg daarbij uit hoe die kenmerken zijn berekend uit de antwoorden op de vragen.
- Om een steekproef te kunnen trekken uit de doelpopulatie hebt u een steekproefkader nodig. Legt uit welk *steekproefkader* u hebt gebruikt. Geef aan of het steekproefkader actueel was. Leg uit of u te maken heeft gehad met zaken als onderdekking en overdekking.
- Beschrijf het *steekproefontwerp*. Hoe hebt u de steekproef getrokken? Was het een steekproef met gelijke of ongelijke kansen? Hebt u de steekproef met of zonder teruglegging getrokken? Hoe groot waren die kansen precies?
- Beschrijf hoe het *veldwerk* van de peiling is uitgevoerd. Was het een mondelinge (face-to-face), telefonische of schriftelijke enquête? Is gebruik gemaakt van een papieren of elektronische vragenlijst? Zijn er interviewers ingeschakeld? Waren dit ervaren interviewers? Hebben ze nog speciale training gehad? Kregen de interviewers nog te maken met bijzondere problemen?
- Na afloop van het veldwerk controleert u de verzamelde gegevens. Leg uit hoe u de ontdekte fouten hebt gecorrigeerd. Als u *imputatietechnieken* hebt toegepast, legt dan uit hoe welke dat waren en voor welke variabelen ze zijn gebruikt.
- Tijdens het veldwerk voor uw peiling krijgt u te maken met *non-respons*. U moet aangeven hoe hoog de non-respons was. En u moet daarbij de non-respons uitsplitsen naar de verschillende oorzaken (geen contact, weigering en niet in staat).
- Als er sprake is van een flinke hoeveelheid non-respons, dan zult u daarvoor moet corrigeren met een *weegprocedure*. Leg uit welke weegprocedure u hiervoor hebt toegepast. Welke hulpvariabelen hebt u daarvoor gebruikt?
- Uiteindelijk hebt u *schattingen* gemaakt van allerlei populatiekenmerken. Leg uit hoe die schattingen precies in hun werk gingen. Hebt u daarin de correctiegewichten meegenomen? U kunt overwegen de wetenschappelijke formules voor de schattingen in de bijlagen op te nemen.
- Aangezien uw peiling gebaseerd is op een steekproef uit de doelpopulatie, hebben de schattingen *onzekerheidsmarges*. U moet aangeven hoe groot die marges zijn. Als u ook nog te maken hebt met een flinke hoeveelheid non-respons, dan moet u waarschuwen dat er, naast de onzekerheidsmarge van de steekproef, ook nog een vertekening kan zijn.

11.3 Deel 3: de uitkomsten

Het derde deel van het onderzoeksrapport bevat de analyse van de gegevens, Dit deel zou kunnen beginnen met een exploratieve analyse van de gegevens. Daarmee krijgt u inzicht in elke gemeten variabele apart.

U hebt altijd de keuze om de verdeling van een variabele grafisch of numeriek weer te geven. Grafieken zijn vaak eenvoudiger te ‘lezen’ en geven daarom meer inzicht (‘één plaatje zegt meer dan 1000 woorden’). Daarom zou u de voorkeur kunnen geven aan grafieken. Tabellen met numerieke overzichten kunt eventueel opnemen in de bijlagen.

De exploratieve analyse kunt u laten volgen door een diepgaandere analyse waarin u probeert verbanden te leggen tussen variabelen. Ook hier kunt u weer kiezen tussen grafieken en tabellen. De grafieken geven het globale plaatje en de tabellen bevatten de numerieke details.

Bij het beschrijven van de uitkomsten van de analyses moet u niet vergeten te melden dat er sprake is van onzekerheidsmarges. Waar mogelijk moet u die marges opnemen in de tekst.

In de beschrijving van de analyse dient al te veel technische details te vermijden, aangezien dit de leesbaarheid van het verhaal niet ten goede komt. Zo nodig kunt u die technische details opnemen in de bijlagen.

11.4 Deel 4: de conclusies

Het vierde deel van het onderzoeksrapport bevat de conclusies die u uit de peiling hebt getrokken. Het is een terugvertaling van de resultaten van de peiling naar de praktijk. De conclusies staan ook al in deel 1, maar deel 4 kan wat meer een interpretatie zijn. In die zin zouden ze ook wat subjectiever kunnen zijn. Niettemin moet alles wat u zegt niet in tegenspraak zijn met de uitkomsten van de peiling.

Het zou ook kunnen zijn dat uw conclusie een hypothese is over hoe de zaken in elkaar zitten. In dit geval moet er dan een nieuw onderzoek komen om die hypothese te toetsen.

11.5 Deel 5: de literatuur

Het vijfde deel van het onderzoeksrapport bevat een overzicht van de relevant literatuur. Die literatuur valt in feite in twee delen uiteen:

- Inhoudelijke literatuur. Dit zijn publicaties over het onderwerp dat u in de peiling onderzoekt.
- Methodologische literatuur. Dit zijn methodologische publicaties die hebt geraadpleegd voor het op verantwoorde wijze opzetten en uitvoeren van de peiling.

11.6 Deel 6: de bijlagen

In de bijlagen kunt u zaken opnemen die relevant zijn voor de peiling, maar die wellicht te groot of te ingewikkeld zijn voor de lopende tekst van het onderzoeksrapport. Dat soort informatie is echter wel van belang bij het bepalen of het onderzoek wel goed is uitgevoerd.

Zaken die u in de bijlagen kunt opnemen, zijn

- De vragenlijst
- Een exploratieve analyse met voor elke variabele een numeriek overzicht van de verdeling van de antwoorden.
- Formules van de schattingsprocedures.
- Brieven (of emails) die zijn verstuurd aan de respondenten, inclusief herinneringsbrieven (of emails)

12. Een checklist voor peilingen

12.1 Het kaf en het koren

We peilen heel wat af in Nederland. Dat is vooral te merken in de periode voor de Tweede Kamerverkiezingen. In die campagnetijd volgen de politieke peilingen elkaar in hoog tempo op. Maar ook buiten de verkiezingen vragen we steeds vaker de mening van ‘de Nederlander’ over allerlei onderwerpen.

Al die peilingen (enquêtes, surveys) hebben gemeen dat we aan een steekproef van personen uit een populatie een vragenlijst voorleggen. Die vragen kunnen gaan over feitelijke zaken, gedrag en meningen. Met de verkregen antwoorden proberen we dan uitspraken te doen over de gehele populatie. Dat kan, maar dan moet het onderzoek wel op wetenschappelijk verantwoorde wijze zijn opgezet en uitgevoerd.

Voorals het internet is de oorzaak van een sterke toename van het aantal peilingen. Het internet maakt het mogelijk eenvoudig, snel en goedkoop bij heel veel mensen informatie te verzamelen. Er zijn websites waarop iedereen zonder enige kennis van onderzoeksmethoden snel een peiling in elkaar kan timmeren. De vraag is echter of al die peilingen wel goed in elkaar zitten. En als dat niet zo is, dan moet u grote vraagtekens zetten bij de betrouwbaarheid en validiteit van de uitkomsten.

We zien dus steeds meer onderzoek langskomen. Daar zitten goede en slechte onderzoeken bij. Voor de gebruikers van de uitkomsten van peilingen (journalisten, bestuurders, beleidsmakers) is het lastig om op simpele wijze het kaf van het koren te scheiden. Daarom is een checklist gemaakt. Door de vragen in de checklist één voor één af te lopen, krijgt u een eerste indruk van de kwaliteit van een onderzoek. Lijkt de kwaliteit goed te zijn, dan verdienen de uitkomsten van de peiling misschien wel uw nadere aandacht. Roept het doorlopen van de checklist veel vragen op over de kwaliteit van het onderzoek, dan is het misschien maar beter het onderzoek te laten voor wat het is. U kunt echter natuurlijk ook bij uw beschrijving van het onderzoek vermelden dat er problemen zijn, waardoor de conclusies op los zand zijn gebaseerd.

We hebben geprobeerd de checklist simpel te houden. Daardoor heeft hij een wat uitgesproken karakter gekregen: een onderzoek is goed of fout. Uiteraard is de praktijk vaak ingewikkelder, en dan zal het eindoordeel ook wat genuanceerder moeten zijn.

Hieronder lichten we de vragen nog wat nader toe. Daarbij proberen we aan te geven waarom de beantwoording van de vraag relevant is voor de beoordeling van de kwaliteit van het onderzoek.

De checklist was een gezamenlijk initiatief van het Nederlandstalig Platform voor Survey-onderzoek (NPSO), het Centraal Bureau voor de Statistiek (CBS) en de Vereniging voor Onderzoeksjournalisten (VVOJ).

12.2 De checklist

1.	<p>Is het duidelijk wie de opdrachtgever en/of financier van het onderzoek is? Dan kan worden vastgesteld of die belang heeft bij de uitkomsten. Dat is bijvoorbeeld het geval als het onderzoek wordt uitgevoerd in het kader van de promotie van een product, dienst of standpunt.</p> <ul style="list-style-type: none"> • Ja: Ga door naar 2. • Nee: Let op! Er bestaat een risico dat de objectiviteit van het onderzoek niet is gegarandeerd.
2.	<p>Is er een onderzoeksverantwoording waarin precies staat aangegeven hoe het onderzoek is opgezet en uitgevoerd?</p> <ul style="list-style-type: none"> • Ja: Ga door naar 3. • Nee: Let op! De betrouwbaarheid van het onderzoek kan niet worden vastgesteld.
3.	<p>Is duidelijk wat de doelpopulatie is? Dit is de groep die is onderzocht en waarop de conclusies van het onderzoek betrekking hebben.</p> <ul style="list-style-type: none"> • Ja: Ga door naar 4. • Nee: Let op! De uitkomsten kunnen niet in de juiste context worden geïnterpreteerd.
4.	<p>Om de kwaliteit van de vragenlijst te kunnen beoordelen, moet in ieder geval zijn voldaan aan de volgende twee voorwaarden:</p> <ul style="list-style-type: none"> ○ De volledige vragenlijst is opgenomen in de onderzoeksverantwoording; ○ De vragenlijst is voor de start van het onderzoek getest. <p>Is aan deze voorwaarden voldaan?</p> <ul style="list-style-type: none"> • Ja: Ga door naar 5. • Nee: Let op! De uitkomsten van het onderzoek kunnen onbetrouwbaar zijn
5.	<p>Hoe is de steekproef getrokken? Is de steekproef geloot met een kans-steekproef waarin elke persoon in de doelgroep een positieve kans had om in de steekproef te komen? Die kansen moeten bij voorkeur gelijk zijn. In ieder geval moeten we de kansen altijd kunnen worden rekenen.</p> <ul style="list-style-type: none"> • Geloot uit de hele groep. Ga door naar 6. • Geloot uit deel van de groep. Bijvoorbeeld alleen uit de internetbezitters of alleen uit personen die in het telefoonboek staan. Ga door naar 6, maar besef dat de uitkomsten betrekking hebben op een kleinere groep. • Zelfselectie via internet. Let op! Er zijn geen betrouwbare uitspraken mogelijk. • Een andere vorm van selectie met onbekenden trekkingskansen, bijvoorbeeld een quota-steekproef. Let op! Er zijn geen betrouwbare uitspraken mogelijk.
6.	<p>Is de omvang van de gerealiseerde steekproef vermeld? Het gaat hier om het aantal respondenten.</p> <ul style="list-style-type: none"> • Ja: Ga door naar 7. • Nee: Let op! Je kunt de onzekerheidsmarges van de uitkomsten niet vaststellen.
7.	<p>Is het percentage respons voldoende hoog, zeg hoger dan 50%?</p> <ul style="list-style-type: none"> • Ja: Ga door naar 8. • Nee: Let op! Een lage respons kan leiden tot een grote mate van selectiviteit in het onderzoek en dus tot onjuiste uitkomsten.

8.	<p>Is een correctie (weging) uitgevoerd voor de opgetreden non-respons?</p> <ul style="list-style-type: none"> • Ja: Ga door naar 9. • Nee: Let op! Non-respons leidt vaak tot een vertekening in de uitkomsten.
9.	<p>Zijn de onzekerheidsmarges van de uitkomsten vermeld? Dit is de afwijking die wordt veroorzaakt door dat niet iedereen in de steekproef zit, maar slechts een selectie.</p> <ul style="list-style-type: none"> • Ja. Merk op dat in die marges niet de vertekening ten gevolg van non-respons en eventuele andere effecten (bijvoorbeeld geheugeneffecten) zijn meegenomen. De onzekerheid kan dus nog groter zijn. • Nee. Let op! Het is dan lastig om de uitkomsten op hun juiste waarde te schatten. Je kunt echte effecten niet onderscheiden van de ‘ruis’ van de steekproef.

12.3 Toelichting

12.3.1 *Wie is de opdrachtgever?*

Het is van belang te weten wie de opdrachtgever en/of financier is van het onderzoek. Het komt voor dat die er belang bij heeft dat de uitkomsten een bepaalde richting uitgaan. Zo zie je regelmatig persberichten met de uitkomsten van onderzoek waaruit zou moeten blijken dat bepaalde producten of diensten erg goed zijn. Die persberichten blijken dan afkomstig te zijn van bedrijven die deze producten of diensten zelf leveren. Het gaat dan louter om de publiciteit en niet om objectief onderzoek.

De BBC, bijvoorbeeld, kent een richtlijn dat de opdrachtgever van de peiling de uitkomsten van een opiniepeiling niet in de uitzending mag komen toelichten of interpreteren. De omroep dient zich te beperken tot het slechts melden van de uitkomsten. Ook moeten de programmamakers de opdrachtgever vermelden. Uiteraard moet dan wel bekend zijn wie de opdrachtgever is.

12.3.2 *Is er een onderzoeksverantwoording beschikbaar?*

De onderzoeksverantwoording is een rapport dat precies beschrijft hoe het onderzoek is opgezet en uitgevoerd. Het rapport moet voldoende informatie bevatten om te kunnen vaststellen of het om een wetenschappelijk verantwoorde onderzoek gaat.

In de onderzoeksverantwoording moet minimaal de volgende aspecten aan de orde komen:

- De naam van de opdrachtgever.
- De naam van de organisatie die de peiling heeft uitgevoerd.
- De doelpopulatie. Dat is de groep waarop de uitkomsten van het onderzoek betrekking heeft.
- De volledige vragenlijst. Het moet duidelijk of de vragenlijst ook van te voren is getest.
- Het gebruikte steekproefkader. Dit is de lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan (adres, telefoonnummer, e-mailadres, e.d.).
- De wijze waarop de steekproef is getrokken. De verantwoording moet aangeven of er is geloot en hoe er is geloot.
- De omvang de bruto steekproef. Dit is de omvang van de steekproef zoals hij is getrokken uit het steekproefkader.
- De omvang van de respons (netto steekproef). Dit is het aantal respondenten.
- Het percentage respons ($100 \times \text{Respons} / \text{Bruto steekproef}$).

- De wijze waarop de peiling eventueel is gecorrigeerd voor non-respons (en mogelijke andere selectie-effecten). Welke weegvariabelen zijn gebruikt? En hoe zijn ze gebruikt voor het berekenen van de gewichten?
- De omvang van de onzekerheidsmarges. Merk op dat je die alleen kunt berekenen als de steekproef is geloot, en de non-respons niet tot vertekening leidt. Ingeval van aanzienlijke non-respons of zelfselectie kunnen de afwijkingen nog veel groter zijn. Is deze waarschuwing opgenomen?

12.3.3 Wat is de doelpopulatie

De *doelpopulatie* is de groep van mensen waarop de uitkomsten van het onderzoek betrekking heeft. De doelpopulatie moet duidelijk zijn omschreven.

Problemen kunnen ontstaan als het steekproefkader niet de hele populatie afdekt. In feite trek je de steekproef dan uit een andere, kleinere groep. De consequentie daarvan is dat de uitkomsten dan ook betrekking hebben op die kleinere groep en niet op de oorspronkelijke populatie.

Voorbeeld: U definieert de doelpopulatie als alle Nederlanders van 18 jaar en ouder, terwijl u de steekproef trekt uit bezitters van een internetverbinding thuis. De conclusies van het onderzoek hebben dan alleen betrekking op de internetbezitters van 18 jaar en ouder.

12.3.4 Hoe zit het met de kwaliteit van de vragenlijst?

Een goede *vragenlijst* is van cruciaal belang. Het is niet zo moeilijk de uitkomsten van het onderzoek te beïnvloeden door te manipuleren met de teksten van de vragen en de volgorde van de vragen. In een goede vragenlijst zijn de vragen objectief en begrijpelijk geformuleerd. In ieder geval moet je de volgende valkuilen vermijden:

- Onbegrijpelijke vragen. Vragen kunnen onbegrijpelijk zijn voor de respondenten door jargon, ander onbekend woordgebruik, of door lange, vage of ingewikkelde zinsconstructies. Voorbeeld: *Bent u tevreden over de recreatieve voorzieningen in uw woonplaats?*
- Vragen die je op verschillende manieren kunt interpreteren. Voorbeeld: *Wanneer ging u van school af?* Gaat het hier om een datum, leeftijd of andere gebeurtenis (toen ik in verwachting was)?
- Suggestieve en sturende vragen. Voorbeeld: *Vindt u ook niet dat ... ?*
- Vragen waarin je meer dan één vraag stelt. Voorbeeld: *Heeft u deze baan voor korte tijd om iets bij te verdienen?*
- Vragen die een ontkenning of dubbele ontkenning bevatten. Voorbeeld: *Gebruikt u liever niet een niet-medicinale shampoo?*
- Vragen die teveel een beroep doen op de herinnering. Dit levert vooral problemen op als het om een relatief lange periode gaat en je vraagt naar minder

belangrijke gebeurtenissen die vrij regelmatig voorkomen. Voorbeeld: *Hoe vaak hebt u in het afgelopen jaar een bezoek gebracht aan de supermarkt?*

Voor een goed onderzoek is het noodzakelijk dat je de vragenlijst test voordat hij 'het veld' in gaat.

12.3.5 *Hoe is de steekproef getrokken?*

Om op verantwoorde wijze conclusies te kunnen trekken over de doelpopulatie, moet u de steekproef door loting trekken. Het moet een *kanssteekproef* zijn. Iedereen in de populatie moet een positieve kans hebben gehad om in de steekproef te komen. De trekkingskansen moeten bekend zijn.

De eenvoudigste wijze van trekken is die waarbij iedereen dezelfde kans heeft gehad om in de steekproef te komen. Dan geldt het principe dat het percentage (of gemiddelde) in de steekproef een goede schatter is voor het percentage (of gemiddelde) in de populatie.

Het is denkbaar om de steekproef met ongelijke kansen te trekken. Het berekenen van schattingen voor de doelpopulatie is dan wat ingewikkelder, omdat u moet corrigeren voor die ongelijke kansen. Een voorbeeld is een steekproef van personen die wordt verkregen door eerst adressen met gelijke kansen te loten en vervolgens op elk geselecteerd adres een willekeurig persoon te kiezen. Personen in grote gezinnen hebben dan een kleinere trekkingskans dan personen in kleine gezinnen.

U trekt de steekproef uit een *steekproefkader*. Dit is de (elektronische) lijst waarin voor ieder lid van de doelpopulatie de contactgegevens staan (adres, telefoonnummer, e-mailadres, e.d.). Als het steekproefkader niet de gehele doelpopulatie omvat, dan hebben de conclusies van het onderzoek alleen betrekking op het deel van de doelpopulatie dat u via het steekproefkader kunt bereiken.

Als u de steekproef niet hebt verkregen door loting, dan zijn de trekkingskansen onbekend en kunt u nooit goede schattingen voor populatiepercentages (en populatiegemiddelden) berekenen. Dit is bijvoorbeeld het geval bij peilingen via het internet waarbij u de steekproef hebt verkregen via zelfselectie.

12.3.6 *Hoe groot is de steekproef?*

Als de je steekproef door loting hebt verkregen, dan kun je de precisie van de schattingen berekenen. Daarbij geldt de eenvoudige regel dat de precisie toeneemt als de steekproef groter is.

We geven de precisie meestal weer in de vorm van een onzekerheidsmarge. Die *onzekerheidsmarge* geeft aan hoeveel de schatting maximaal kan afwijken van de werkelijke waarde in de doelpopulatie. Tabel 12.1 bevat voor verschillende steekproefgroottes en een reeks van steekproefpercentages de bijbehorende onzekerheidsmarge.

Tabel 12.1. Onzekerheidsmarges

Percentage	Omvang van de steekproef					
	100	200	500	1000	2000	5000
10	5,9	4,2	2,6	1,9	1,3	0,8
20	7,9	5,6	3,5	2,5	1,8	1,1
30	9,0	6,4	4,0	2,8	2,0	1,3
40	9,7	6,8	4,3	3,0	2,1	1,4
50	9,8	6,9	4,4	3,1	2,2	1,4
60	9,7	6,8	4,3	3,0	2,1	1,4
70	9,0	6,4	4,0	2,8	2,0	1,3
80	7,9	5,6	3,5	2,5	1,8	1,1
90	5,9	4,2	2,6	1,9	1,3	0,8

Stel u hebt voor een peiling een steekproef van 500 personen getrokken. Verder blijkt dat 40% van de respondenten voorstander is van een bepaalde maatregel is. Bij die 500 en die 40% hoort een marge van 4,3%. Het percentage voorstanders in de doelpopulatie zal dus liggen tussen $40 - 4,3 = 35,7\%$ en de $40 + 4,3 = 44,3\%$.

Stel dat u in een peiling onder 1.000 personen vraagt naar de politieke voorkeur. Uit de peiling blijkt dat 20% op een bepaalde partij gaat stemmen. Een maand later voert u een nieuwe peiling uit, ook onder 1.000 personen. Nu blijkt 22% op die partij te gaan stemmen. Kunt u nu zeggen dat de aanhang van die partij is toegenomen? Nee, want beide percentage hebben een onzekerheidsmarge van 2,5%. Die marge is groter dan het verschil tussen de percentage stemmers ($22 - 20 = 2\%$). Het verschil kan dus net zo goed veroorzaakt door de 'ruis' in de steekproef.

12.3.7 Hoe groot is de non-respons?

Non-respons is het verschijnsel dat in de steekproef getrokken personen niet de gewenste gegevens verstrekken. Het lukt niet contact met ze te leggen (bijvoorbeeld omdat ze zelden thuis zijn), ze niet in staat zijn om de vragen te beantwoorden (door bijvoorbeeld ziekte of taalproblemen) of ze weigeren mee te doen. Non-respons leidt er heel vaak toe dat bepaalde groepen mensen over- of ondervetegenwoordigd zijn in de steekproef. De respons is dan selectief. Daardoor is de gerealiseerde steekproef niet meer representatief. Er zijn drie factoren die de omvang van de vertekening in de uitkomsten bepalen:

- Hoeveel mensen doen niet mee? Is de respons bij een peiling laag, dan is het risico groot dat de uitkomsten ernstige afwijkingen vertonen.
- Zijn non-respondenten anders? Bijvoorbeeld: bij verkiezingsonderzoek zien we steeds weer een sterke samenhang tussen respons en stemgedrag. Mensen die niet meedoen aan het onderzoek, gaan ook niet stemmen. En respondenten gaan juist wel stemmen. Dit leidt ertoe dat stemmers zijn oververtegenwoordigd in de peiling. Als gevolg daarvan schat u de opkomstpercentages te hoog.
- Doen sommige mensen vaker aan onderzoek mee dan anderen? Als er mensen zijn met hoge responskansen (ze doen vaak mee aan dit soort onderzoek) en mensen met lage responskansen (ze doen bijna nooit mee aan dit soort onderzoek), dan verhoogt dit de vertekening in de uitkomsten. Als iedereen even geneigd is om aan het onderzoek mee te doen (hun responskansen zijn gelijk), dan leidt non-respons niet tot een vertekening.

U kunt meestal niet vaststellen hoe groot de vertekening is. Dat kunt u alleen als bekend is wat de antwoorden van de non-respondenten op de vragen zouden zijn. Omdat het echter om non-respondenten gaat, zijn die antwoorden dus onbekend. Het is wel mogelijk om uit te rekenen hoe groot de afwijking ten gevolge van non-respons maximaal kan zijn.

Stel eens dat slechts 40% van de mensen in de steekproef respondeert in een verkiezingsonderzoek. Van die respondenten zegt 60% te zullen gaan stemmen. Als 40% respondeert, dan respondeert 60% dus niet. Stel dat in een extreme situatie al die non-respondenten niet gaan stemmen. Dan is het percentage stemmers in de hele steekproef

$$0,40 \times 60\% + 0,60 \times 0\% = 24\%.$$

In de andere extreme situatie gaan alle non-respondenten wel stemmen. Dan is het percentage stemmers in de hele steekproef

$$0,40 \times 60\% + 0,60 \times 100\% = 84\%.$$

Dus het in de respons gevonden percentage stemmers van 40% had ook 24% kunnen zijn en het had ook 84% kunnen zijn. Het werkelijk percentage stemmers zal ergens tussen 24% en 84% liggen. Een dergelijke grote bandbreedte maakt het onderzoek onbruikbaar. Gelukkig doen die extreme situaties zich in de praktijk niet vaak voor, maar het effect van non-respons kan wel groot zijn!

Tabel 12.2 bevat de grenzen voor de werkelijke percentages in de steekproef voor een reeks aan verschillende responspercentages. Duidelijk is te zien dat de omvang van de maximale afwijking kleiner wordt naarmate het percentage respons groter is.

Tabel 12.2. Bandbreedte ten gevolge van non-respons

Gevonden percentage	Percentage respons			
	20	40	60	80
10	2 – 82	4 – 64	6 – 46	8 – 28
20	4 – 84	8 – 68	12 – 52	16 – 36
30	6 – 86	12 – 72	18 – 58	24 – 44
40	8 – 88	16 – 76	24 – 64	32 – 52
50	10 – 90	20 – 80	30 – 70	40 – 60
60	12 – 92	24 – 84	36 – 76	48 – 68
70	14 – 94	28 – 88	42 – 82	56 – 76
80	16 – 96	32 – 92	48 – 88	64 – 84
90	18 – 98	36 – 96	54 – 94	72 – 92

12.3.8 Is er voor non-respons gecorrigeerd?

Het is gebruikelijk om te corrigeren voor non-respons door een weging uit te voeren. Daarbij kent u aan elke respondent een gewicht toe. U berekent die gewichten zo dat ze corrigeren voor de over- of ondervertegenwoordiging van groepen in de respons.

Een voorbeeld: uit een peiling blijkt dat de respondenten voor 60% uit mannen bestaan en voor 40% uit vrouwen. Dat is niet goed want de Nederlandse bevolking bestaat voor 49,5% uit mannen en voor 50,5% uit vrouwen. Kennelijk hebben de mannen beter gerepsondeerd in de peiling want ze zijn oververtegenwoordigd. Om hiervoor te corrigeren, krijgt elke responderende man een gewicht van

$$49,5 / 60,0 = 0,825.$$

Dit betekent dat elke man nog maar meetelt voor 0,825 in plaats van 1. Het gewicht is kleiner dan 1 omdat er teveel mannen waren in de respons. Elk vrouw krijgt een gewicht van

$$50,5 / 40,0 = 1,263.$$

Dus elke vrouw telt mee voor 1,263 in plaats van 1. Het gewicht is groter dan 1 omdat er te weinig vrouwen waren in de respons.

Door het toekennen van gewichten aan mannen en vrouwen maakt u de respons representatief met betrekking tot de variabele geslacht. U kon die gewichten uitrekenen omdat de juiste percentages in de populatie bekend waren. Het idee achter wegen is nu dat u de respons representatief maakt met betrekking tot zoveel mogelijk variabelen. Noodgedwongen kunnen dat alleen variabelen zijn die in de peiling zijn gemeten waarvoor de verdeling in de populatie bekend is. Voorbeelden van veel gebruikte variabelen zijn geslacht, leeftijd, burgerlijke staat en provincie waar men woont.

De hoop is nu dat als u de respons representatief maakt met betrekking tot een heleboel variabelen, hij misschien ook wel representatief wordt met betrekking tot de verschijnselen die u onderzoekt. Daarmee zou dan de vertekening in de uitkomsten verdwijnen.

Helaas is niet elke weging effectief. Wegen kan alleen een vertekening verminderen als voldaan is aan twee voorwaarden:

- Er is een sterk verband tussen de variabelen die u onderzoekt (de doelvariabelen) en de variabelen die u gebruikt voor wegen (de hulpvariabelen).
- Er is een sterk verband tussen het responsgedrag en de hulpvariabelen die u gebruikt voor wegen.

Soms heeft wegen veel effect, bijvoorbeeld als u bij een onderzoek naar tijdbesteding weegt naar het hebben van een baan en de aanwezigheid van kinderen in het gezin. Die factoren hebben ook een invloed op de tijdbesteding. Maar het wegen met leeftijd en geslacht in een opiniepeiling kan zinloos zijn als die variabelen niet de opinie bepalen. Het is in ieder geval van belang om vast te stellen welke groepen ondervetegenwoordigd zijn, wat dit voor effect op de resultaten kan hebben en of wegen de uitkomsten van het onderzoek verbetert.

12.3.9 Zijn er onzekerheidsmarges gegeven?

De uitkomsten van een peiling zijn slechts schattingen van wat er in de populatie aan de hand is. Het is daarom onrealistisch te veronderstellen dat die uitkomsten exact overeenkomen met de werkelijk cijfers voor de hele populatie. Zelfs in het ideale geval van een nette kanssteekproef zonder non-respons is er een afwijking tussen schatting en werkelijke waarde. U kunt uitrekenen hoe groot die afwijking maximaal kan zijn. Dat is de onzekerheidsmarge. Voor het bepalen van de onzekerheidsmarge kunt u tabel 9.1 gebruiken.

Voorbeeld: voor een peiling loot u een steekproef van 500 personen. Iedereen doet mee. Er is geen non-respons. Van de respondenten zegt 60% voorstander te zijn van een bepaalde maatregel. De bijbehorende onzekerheidsmarge is 4,3%. Dan zal het percentage voorstanders in de populatie liggen tussen de $60 - 4,3 = 55,7\%$ en de $60 + 4,3 = 64,3\%$.

Zoals we al eerder hebben aangegeven, kunnen de onzekerheidsmarges als gevolg van non-respons veel groter zijn dan die gebaseerd op de steekproefomvang. De marge die u kunt uitrekenen, is dus in feite een ondergrens. De werkelijke marge kan veel groter zijn.

Als veranderingen in de tijd kleiner zijn dan de onzekerheidsmarges veroorzaakt door de steekproeftrekking, dan is er geen sprake van 'echte' verschillen tussen peilingen. Het heeft dan ook geen zin om te proberen die verschillen te verklaren. En als er wel sprake lijkt te zijn van 'echte' verschillen, kan het optreden van non-respons nog roet in het eten gooien.

12.4 Een voorbeeld

Op 20 februari 2012 trad Job Cohen af als partijleider van de PvdA. Daarmee kwam de weg vrij voor een nieuwe politiek leider en fractievoorzitter voor de partij. Tweede Kamerleden van de PvdA konden zich melden om als fractievoorzitter in elk geval de komende tijd ook politiek leider te zijn.

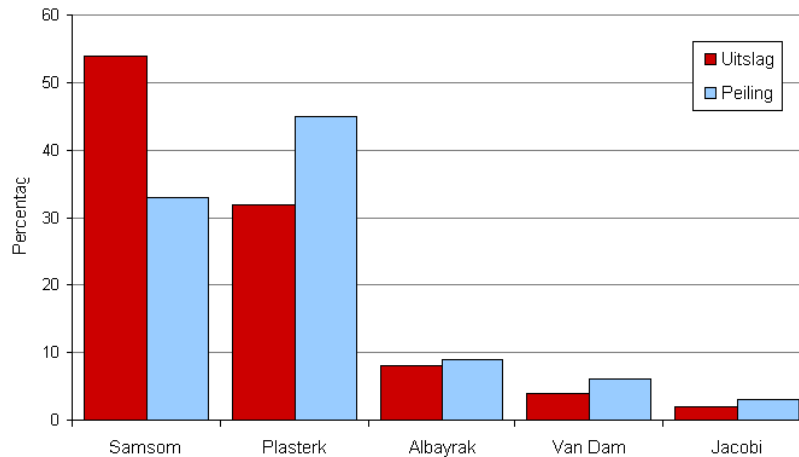
PvdA-partijvoorzitter Hans Spekman vond de mening van de leden van de partij belangrijk. Daarom werd een ledenraadpleging georganiseerd. De uitslag van die ledenraadpleging gold als 'zwaarwegend advies' voor de fractie.

Er waren vijf fractieleden die zich kandidaat stelden voor het voorzitterschap. Dat waren Nebahat Albayrak, Martijn van Dam, Lutz Jacobi, Ronald Plasterk en Diederik Samsom.

Op één dag voor de afsluiting van de ledenraadpleging was er een peiling uit om te zien hoe de vijf kandidaten ervoor stonden. Volgens de peiling zou Ronald Plasterk als winnaar uit de bus komen met 45% van de stemmen. Diederik Samsom zou blijven steken op 33% van de stemmen.

Dit bleek niet juist te zijn. Een dag later was Samsom de grote winnaar met 54% van de stemmen. En Plasterk kreeg slechts 32% van de stemmen. De verschillen tussen de peiling en de werkelijke uitslag waren groot. Dat wordt nog een geïllustreerd in grafiek 12.1. Als je dit zo ziet kun je je afvragen wat zo'n peiling nu eigenlijk waard is. Zet die peiling de mensen niet op het verkeerde been? Het is daarom goed om de checklist eens op deze peiling los te laten.

Figuur 12.1. De peiling en de werkelijke uitslag



12.4.1 Wie is de opdrachtgever?

Uit de beschikbare informatie viel niet af te leiden wie opdracht had gegeven tot het uitvoeren van de peiling. Het leek erop dat het een eigen initiatief was van de peiler. Er was daarom geen sprake van een bedrijf of organisatie die een product of dienst wilde promoten.

12.4.2 Is er een onderzoeksverantwoording beschikbaar?

Er was geen onderzoeksverantwoording beschikbaar. Uit de informatie die op het internet beschikbaar was, kreeg je echter wel een beeld hoe de peiling ongeveer in elkaar zat.

Uitgangspunt was de vorming van een online-panel met daarin zoveel mogelijk mensen die aan dit soort onderzoek willen meedoen. Er was geen steekproef getrokken. Mensen gaven zich zelf op voor het panel. Uit dit panel werd vervolgens voor elke peiling een steekproef getrokken, en wel zo dat die steekproef representatief was met betrekking tot een aantal hulpvariabelen. Omdat er non-respons kon optreden in de peiling, werd achteraf nog een weegprocedure uitgevoerd

12.4.3 Wat is de doelpopulatie?

Ten aanzien van de doelpopulatie was er een groot probleem. Doel van de peiling was inzicht te krijgen in de populariteit van de vijf kandidaten voor het voorzitterschap van de fractie. De leden van de PvdA gingen daarover stemmen. Je zou de steekproef dus moet trekken uit de leden van het PvdA. Dat was bij de peiling niet gebeurd. De steekproef was getrokken uit alle leden van het online-panel die aangaven de PvdA een kans te geven op hun stem. Dat is dus een heel andere populatie. Dit kan voor een flink deel de verschillen tussen peiling en werkelijke uitslag in figuur 9.1 verklaren.

12.4.4 Hoe zit het met de kwaliteit van de vragenlijst?

De vragenlijst voor deze peiling was niet beschikbaar op het internet. Ook was niet duidelijk of de vragenlijst was getest. Daarom valt er ook weinig te zeggen over de kwaliteit ervan.

Je kunt je wel afvragen hoe precies is bepaald of iemand een potentiële PvdA-stemmer is. Wat voor de vraag is daarvoor gesteld? De website van de peiling lijkt te suggereren dat de respondenten is gevraagd of ze de PvdA een kans geven op hun stemmen. Dit soort vragen is lastig, want het zijn hypothetische vragen. Kleine nuanceverschillen kunnen al gauw leiden tot andere antwoorden

12.4.5 Hoe is de steekproef getrokken?

Door het ontbreken van een onderzoeksverantwoording is het niet duidelijk hoe de steekproef voor deze peiling precies is getrokken. De schaarse informatie op het internet lijkt erop te wijzen dat een steekproef is getrokken uit een online-panel. Dat panel is door zelfselectie tot stand gekomen en niet door het trekken van een aselechte steekproef uit de bevolking. Het panel is dus zeker niet representatief.

De steekproef uit het panel is zo getrokken dat die representatief is met betrekking tot een aantal hulpvariabelen. Die steekproef lijkt representatief te zijn, maar omdat hij is getrokken uit een niet-representatief panel, is dat toch niet het geval. Wat niet in het panel zit, krijg je er ook niet in door het op deze manier trekken van de steekproef.

12.4.6 Hoe groot is de steekproef?

Over de omvang van de steekproef wordt vrijwel niets gemeld. Tijdens de verkiezingscampagne voor de nieuwe fractievoorzitter van het PvdA is de peiling vijf keer uitgevoerd. Bij een van die peilingen meldt de website van de peiling dat de uitkomsten zijn gebaseerd op meer dan 750 respondenten. Mogelijk ligt de steekproefomvang voor de andere vier peilingen ook rond die 750. Dat is een steekproef van beperkte omvang. We moeten dus rekening houden met flinke onzekerheidsmarges.

12.4.7 Hoe groot is de non-respons?

Over non-respons in de peiling wordt niets gemeld. De eerder genoemde respons van 750 mensen is wat er aan respondenten is overgebleven na het optreden van non-response (de netto steekproef). Alleen als je ook de initiële omvang van de steekproef (de bruto steekproef) kent, kun je het percentage non-respons berekeningen.

Het is jammer dat het percentage non-respons ontbreekt, want dit cijfer is een belangrijke indicator voor de kwaliteit van de uitkomsten van de peiling. Zeker als de non-respons omvangrijk is, moeten we met ernstige vertekening in de uitkomsten rekening houden.

12.4.8 *Is er voor non-respons gecorrigeerd?*

Er is een weegprocedure toegepast om de uitkomsten te corrigeren voor non-respons. Het is niet duidelijk welke variabelen hiervoor zijn gebruikt. Al eerder hebben we in deze leidraad aangegeven dat een weegprocedure alleen effectief is als de gebruikte weegvariabelen samenhangen met de doelvariabelen van de peiling en het responsgedrag. Door gebrek aan informatie kunnen we dus niet vaststellen of de weging de door non-respons veroorzaakte vertekening kan reduceren.

De website van de peiling meldt één speciale weegvariabele wel. Dat is de variabele die vastlegt op welke partij de persoon heeft gestemd bij de laatste verkiezingen voor de Tweede Kamer. In principe is dat een nuttige weegvariabele omdat hij de peiling representatief maakt met betrekking tot de partijvoorkeur. Dat is belangrijk als je peiling gaat over politieke onderwerpen. Bij deze specifieke peiling ging het echter alleen om de aanhang van de PvdA. Wegen naar partijvoorkeur heeft daarom geen effect op de uitkomsten. Een tweede probleem is dat nieuwe leden van het panel moeten aangeven op welke partij ze hebben gestemd bij de laatste verkiezingen voor de Tweede Kamer. Weten mensen dat nog wel? De partijtrouw is niet meer zo groot als vroeger. Er zijn heel veel zwevende kiezers die pas op het laatste moment hun keuze bepalen. En misschien willen sommige kiezers niet zeggen op welke partij ze hebben gestemd. Kortom het antwoord op deze vraag kan meetfouten bevatten, met alle gevolgen van dien voor het effectiviteit van de weegprocedure.

12.4.9 *Zijn er onzekerheidsmarges gegeven?*

Bij de peiling worden geen onzekerheidsmarges vermeld. Daarom is het onduidelijk of aangetroffen verschillen ook inderdaad ‘echte’ verschillen zijn, of dat we ze kunnen toeschrijven aan de ‘ruis’ van de steekproef. Dat kan ertoe leiden dat de gebruikers van de uitkomsten (bijvoorbeeld de media) inhoudelijke discussies starten over effecten die er helemaal niet zijn.

In het geval van de peiling over het fractievoorzitterschap van de PvdA kunnen we een poging doen de onzekerheidsmarge alsnog te berekenen. De formule voor de (geschatte) onzekerheidsmarge van een gevonden percentage p is bij benadering gelijk aan

$$M = 1,96 \times \sqrt{\frac{p \times (100 - p)}{n}}.$$

We moeten de benaderingsformule gebruiken omdat de originele formule de omvang van de populatie N bevat. Die kennen we niet. We weten niet hoeveel potentiële PvdA-stemmers er in Nederland zijn. Verder nemen we aan dat de steekproefomvang gelijk is aan $n = 750$.

In de peiling had 45% van de respondenten een voorkeur voor Ronald Plasterk. De onzekerheidsmarge is in dit geval

$$M = 1,96 \times \sqrt{\frac{45 \times 55}{750}} = 3,6.$$

Dus zal het populatiepercentage (met 95% zekerheid) liggen ergens tussen de $45 - 3,6 = 42,4\%$ en de $45 + 3,6 = 48,6\%$. Het aangeven van deze bandbreedte geeft een eerlijker beeld van de uitkomst dan het wekken van een schijn van exactheid door slechts het vermelden van één cijfer (45%).

Op dezelfde manier kunnen we uitrekenen dat het percentage voor Samsom zal liggen tussen de 29,6% en de 36,4%.

We zijn er zo nog niet met de onzekerheid in de uitkomsten. Er kunnen nog twee problemen die kunnen leiden tot een vertekening in de uitkomsten. In de eerste plaats zijn de gegevens afkomstig uit een online-panel dat door zelfselectie is gevuld. En in de tweede plaats kan non-respons nog een vertekening veroorzaken. Dit soort vertekeningen kan aanzienlijk zijn. Ze zijn vaak groter dan de onzekerheidsmarges. Ook is het zo dat vergroting van de omvang van de steekproef niet helpt. Een grotere steekproef verkleint wel de steekproefmarge maar niet de vertekening.

12.4.10 Conclusies

Aangezien een toereikende verantwoording van de opzet en uitvoering ontbreekt, kunnen we maar heel weinig zeggen over de betrouwbaarheid en validiteit van de uitkomsten ervan.

Een ander belangrijk probleem is dat de respondenten afkomstig lijken te zijn uit een online-peiling waarvoor de deelnemers zichzelf spontaan hebben aangemeld. De steekproef is niet tot stand gekomen via loting uit de populatie. Daardoor is het niet mogelijk om op een wetenschappelijk verantwoorde manier conclusies te trekken over de doelpopulatie.

Literatuur

- Bethlehem, J.G. (2009a), *Applied Survey Methods – A Statistical Approach*. John Wiley & Sons, Hoboken, NJ.
- Bethlehem, J.G. (2009b), *The Rise of Survey Sampling*. Discussion Paper 09015, Centraal Bureau voor de Statistiek, Den Haag/Heerlen.
- Bethlehem, J.G. & Biffignandi, S. (2012), *Handbook of Web Surveys*. John Wiley & Sons, Hoboken, NJ.
- Bethlehem, J.G., Cobben, F. & Schouten, B. (2011), *Handbook of Nonresponse in Household Surveys*. John Wiley & Sons, Hoboken, NJ.
- Bowley, A.L. (1906), Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science. *Journal of the Royal Statistical Society* 69, blz. 548-557.
- Bowley, A.L. (1926): Measurement of the Precision Attained in Sampling. *Bulletin of the International Statistical Institute*, XII, Book 1, blz. 6-62.
- CBS (1948), Enige beschouwingen over steekproeven. Herdruk uit: *Statistische en Economische Onderzoekingen* 3, Centraal Bureau voor de Statistiek, Den Haag.
- CBS (1987), *CBS Select 4, Automation in Survey Processing*. Centraal Bureau voor de Statistiek, Voorburg.
- Couper, M.P. (2008), *Designing Effective Web Surveys*. Cambridge University Press, New York, USA.
- Couper, M.P., Baker, R.P., Bethlehem, J.G., Clark, C.Z.F., Martin, J., Nicholls II, W.L., O'Reilly, J.M. (red.) (1998), *Computer Assisted Survey Information Collection*. Wiley, New York.
- Den Dulk, C.J. & Van Maarseveen, J.G.S.J. (1990), Volkstellingen 2795-1971. De ontwikkeling van beleid en methode van onderzoek. In: Erwich, B. & Van Maarsseveen, J.G.S.J. (red.), *Een eeuw statistieken*, Centraal Bureau voor de Statistiek, Voorburg/Heerlen, blz. 329-366.
- Dillman D., Smyth J., Christian L. M., (2009), *Internet, Mail and Mixed mode Surveys. The Tailored Design Method*, John Wiley & Sons, Hoboken, NJ, USA.
- Idenburg, Ph. J. (1952), *Steekproeven*. Intern CBS rapport S.708-CC-52. Centraal Bureau voor de Statistiek, Den Haag.
- Kiaer, A. N. (1895), Observations et Expériences Concernant des Dénombrements Représentatives. *Bulletin of the International Statistical Institute*, IX, Book 2, blz. 176-183.

- Kiaer, A. N. (1997, herdruk): Den Repräsentative Undersökelsesmetode. *Christiania Videnskabselskabets Skrifter. II. Historiskfilosofiske klasse*, Nr 4 (1897). Statistisk Sentralbyrå, Oslo, Noorwegen.
- Neyman, J. (1934), On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society* 97, blz. 558-606.
- NIPO (1946a), Wat denkt het publiek ervan? *De Publieke Opinie*, 1^e jaargang, No. 1, blz. 1-2.
- NIPO (1946b), Rekening en Verantwoording? *De Publieke Opinie*, 1e jaargang, No.2, blz. 1.
- Saris, W.E. (1998), Ten years of interviewing without interviewers: the Telepanel. In: Couper, M.P., Baker, R.P., Bethlehem, J.G., Clark, C.Z.F., Martin, J., Nicholls II, W.L., O'Reilly, J.M. (red.) (1998), *Computer Assisted Survey Information Collection*. Wiley, New York, blz. 409-430.
- Tiemijer, W.L. (2008), *Wat 93,7% procent van de Nederlanders moet weten over opiniepeilingen*. Aksant, Amsterdam.

Bronvermelding

Figuur 2.2, *Het Domesday Book*: Andrews, W. (1900), *Historic Byways and Highways of Old Engeland*.

Figuur 2.3, *De Quipucamayoc*: Burland, C.A. (1971), *Peru onder de Inca's*, Thieme, Zutphen.

Figuur 2.4, *De volksteller komt langs*: United States Library of Congress.

Figuur 2.5, *De kranten verklaren Thomas Dewey tot winnaar*: Associated Press / Byron Rollins.

Figuur 6.1, *Een Hollerith-machine*: United States Census Bureau.

Register

A-nummer 62
Aangepaste populatie variantie 28, 29
Aankruisvakje 75
Aanschrijfbrief 68
Adressenbestand 63
Adressensteekproef 98, 103
Analogieprincipe 89, 98
Analyse 121
Aselect getal 54
Aselecte imputatie 83
Aselecte steekproef 9
Aselector 53

Begrijpelijke vraag 31
Beroep op herinnering 34
Betrouwbaar 6, 88
Betrouwbaarheid 92
Betrouwbaarheidsinterval 17, 90, 91, 96, 102, 105
Bevolkingsregister 62
Bovengrens van betrouwbaarheidsinterval 92, 101, 105
Box-and-whisker plot 125

Categorische variabele 25
Centraal Bureau voor de Statistiek (CBS) 145
Checklist 145, 146
Chi-kwadraat-grootheid 136
Cirkeldiagram 128, 134
Cognitief interviewen 42
Computer-Assisted Personal Interviewing (CAPI) 9, 70, 77
Computer-Assisted Self Interviewing (CASI) 70
Computer-Assisted Telephone Interviewing (CATI) 9, 69, 77
Computer-Assisted Web Interviewing (CAWI) 71
Computergestuurd enquêteren 9
Continue variabele 26
Controle 80
Controle en correctie 10, 79
Correctie 82
Correctie voor non-respons 117
Correctiegewicht 117
Correlatiecoëfficiënt 130
Cramér's V 136

Discrete variabele 26
Doelpopulatie 7, 23,, 24, 48, 140, 149
Doelvariabele 26, 27, 119
Domesday Book 14
Donor-imputatie 83
Drietrapssteekproef 60
Dubbele vraag 33

Enkelvoudige aselechte steekproef 56
Enquête 5
Evenwichtige steekproef 17
Exploratieve analyse 121

Filtervraag 37, 44
Frequentieverdeling 129

Gallup 19
Geen contact 113
Geheugeneffect 80
Gesloten vraag, één antwoord 36, 74
Gesloten vraag, meer antwoorden 38, 75
Gevoelige vraag 33
Gewicht 98, 103

Histogram 126
Hollerith-machine 69
Homogene groep 119
HTML 74
Hulpvariabele 27, 115
Hypothetische vraag 34

Imputatie 83, 141
Imputatie van het gemiddelde 83
Indicatorvariabele 26
Inductieve analyse 121
Informed consent 45
Insluitkans 61
Integraal onderzoek 13
Internetpeiling 71, 78
Item non-respons 111

Kanssteekproef 20
Keuzelijst 74
Keuzerondje 74
Kruistabel 135

Kwalitatieve variabele 25
Kwantitatieve variabele 26

Literary Digest 19
Loten 17
Luisteronderzoek 24

Matrixvraag 37, 76
Meetfout 80
Methodologische verantwoording 140
Mixed-mode peiling 72
Modus 129
Mondeling enquêteren 8, 67, 68

Nauwkeurige schatter 88
Nederlands Instituut voor de Publieke Opinie (NIPO) 21
Nederlandse Stichting voor de Statistiek (NSS) 21
Nederlandstalig Platform voor Survey-onderzoek (NPSO) 145
Niet in staat 113
Non-respons 10, 79, 111, 141, 151
Normale verdeling 17
NS Publieksprijs 73
Numeriek overzicht 127, 138
Numerieke vraag 39

Omvang van de steekproef 106, 107, 109, 150
Onderdekking 48, 72, 79
Ondergrens van betrouwbaarheidsinterval 92, 101, 105
Onderzoeksrapport 139
Onderzoeksverantwoording 148
Ondubbelzinnige vraag 32
Online enquêteren 9, 67
Ontkenning (dubbele) 33
Onzekerheidsmarge 11, 79, 90, 91, 92, 101, 105, 141, 152
Opdrachtgever 148
Open vraag 35, 75
Opiniepeiling 18
Overdekking 49, 63, 64

Peildatum 7, 23, 24
Peiling 5
Populatiefractie 29
Populatiegemiddelde 28, 29, 93, 103
Populatiekenmerk 8, 23, 28, 141
Populatiepercentage 29, 89

Populatietotaal 28, 29
Populatievariantie (aangepaste) 28, 29
Postcodebestand 48
Precisie 88
Primacy effect 36, 39, 75
Publicatie 11, 139
Puntenwolk 129

Quipu 15
Quipucamayoc 15
Quota-steekproef 20

Random Digit Dialing (RDD) 49, 50
Recency effect 36
Regressielijn 131
Relatiecontrole 81
Relatiefout 81
Representatief 52, 71
Representatieve Methode 17
Respondent 8
Responsgedrag 119
Responspercentage 112, 116, 117
Route-instructie 44
Routecontrole 81
Routefout 82

Samengesteld staafdiagram 132
Samplonië 123
Satisficing 38, 39, 75, 80
Schatting 87
Schatting 11, 87, 99, 141
Schriftelijk enquêteren 8, 67
Selectieve non-respons 10, 111, 115
Snorrendoos 125, 137
Sociaalwenselijk antwoord 33
Spreidingsdiagram 124, 136
Sprongopdracht 41, 44
Staaftdiagram 128
Standaardfout van de schatter 90, 92, 101, 105
Stapeldiagram 133, 134
Staplengte 58
Startgetal 58
Steekproef 5, 52, 150
Steekproef met teruglegging 55
Steekproef zonder teruglegging 55

Steekproeffout 79
Steekproefgemiddelde 93
Steekproefkader 47, 48, 116, 141
Steekproefonderzoek 16, 22
Steekproefontwerp 9, 141
Steekproefpercentage 89
Straight-lining 38, 76
Straw poll 18
Suggestieve vraag 33
Survey-methodologie 7
Survey-onderzoek 5
Synthetische waarde 83
Systematische steekproef 9, 56, 58

Tekstveld 75
Tekstvak 76
Telefonisch enquêteren 8, 67, 68
Telefoonboek 64
Telescoping 34
Testen van vragenlijsten 42
Trekingsdatum 24
Trekingskans 61
Tweetrapssteekproef 9, 56, 60, 97

Uitschieter 121
Unit non-respons 111

Valide 6, 42, 88
Variabele 8, 23, 25, 140
Variantie 74
Variantie van de schatter 91, 95, 101, 105
Veldwerk 10, 67, 141
Vereniging van Onderzoeksjournalisten (VVOJ) 145
Vertekening 111, 115
Verzamelen van gegevens 67
Volgorde van de vragen 40
Volkstelling 14, 15
Vox populi. Vox pop 47
Vraag 31
Vraagtekst 31
Vragenlaboratorium 45
Vragenlijst 6, 8, 31, 141, 149
Vragenmoetheid 43
Vuistregelinterval 127

Waardecontrole 80
Waardefout 81
Weegprocedure 117
Weet niet 32, 71
Wegen 10, 117, 141, 152
Weigering 113

Zelfselectie 72
Zuivere schatter 87