

Inzicht & Impact



Marketing intelligence

Whitepaper

Steekproeven: de basis

Juli 2017, Durk Bosma en Erik de Kort



► Inhoud

Over steekproeven: inleiding	3
Steekproeven: de hoofdzaak	6
De representatieve steekproef	8
De stappen van steekproeftrekking	10
Het issue non-respons	16

1

Over steekproeven: inleiding



Als je een onderzoek gaat opzetten en je hebt besloten dit te doen door middel van het ondervragen van mensen, sta je op een gegeven moment stil bij de vraag welke mensen en hoeveel mensen je wilt ondervragen. Daaraan gekoppeld is de vraag waar je die mensen vindt. We komen aan bij het terrein 'steekproef', één van de lastigste onderwerpen in marktonderzoek. Van Dale beschrijft de steekproef als 'onderzoek van een deel om daarnaar het geheel te kunnen beoordelen'. We noemen het 'geheel' de 'populatie' en het 'deel' de 'steekproef'. De keuzes die je in deze stap maakt, bepalen voor een groot deel de waarde van je onderzoek.

De theorie achter steekproeven zegt het volgende. Als je van een **bepaalde groep mensen** wilt weten hoe ze over iets denken of wat ze doen, hoef je niet de hele groep te ondervragen (of observeren), maar krijg je met het ondervragen van een deel ook een **behoorlijk goed beeld**.

Voorwaarde is wel dat de groep die je ondervraagt voldoende groot is en een **goede afspiegeling** is van de totale groep waarover je iets wilt weten.

Wat precies bedoeld wordt met de vetgedrukte woorden, komen we later op terug. Je moet een aantal keuzes maken om de steekproef te bepalen, onder andere over de gewenste samenstelling van je steekproef en de grootte ervan.

Zes vragen helpen je om de juiste keuzes in deze stap te maken. Deze vragen en hun consequenties behandelen we in deze whitepaper.

1. Hoe ziet de populatie eruit?
2. Op welke kenmerken wil ik dat de steekproef representatief is?
3. Welke vragen moet ik stellen om vast te stellen wat de kwaliteit van mijn steekproef is?
 - Selectievragen/quota
 - Controlevragen
4. Hoe vind ik de relevante mensen?
5. Hoe nauwkeurig en betrouwbaar wil ik dat de resultaten zijn?
6. Hoe kan ik ervoor zorgen dat ze bereid zijn om mee te doen aan mijn onderzoek?

Vogelvlucht

We nemen je eerst mee in een vogelvlucht door de ontwikkelingen in de tijd. De steekproeftrekking is oorspronkelijk gebaseerd op de waarschijnlijkheidsleer zoals die vanaf het einde van de 19e eeuw werd ontwikkeld. Men ontdekte dat het mogelijk was om kwantitatieve uitspraken te doen over een (grote) populatie op basis van metingen bij een kleine subgroep uit deze populatie: de steekproef. Vanuit de theorie van de waarschijnlijkheid was dit alleen mogelijk als er sprake was van enige vorm van willekeurige selectie (ofwel a-selecte trekking), waarbij de kans dat iemand uit de populatie in de steekproef terecht kwam bekend was. We spreken over de kanssteekproef. Bij een 'voldoende' omvang van de steekproef en een a-selecte trekking wordt de steekproef automatisch 'representatief'. Anders gezegd, een verkleinde kopie van de populatie. De waargenomen resultaten in de steekproef mogen dan worden geprojecteerd op de totale populatie. Vanuit deze aanname zijn ook de berekeningen van betrouwbaarheid, significantie en nauwkeurigheid van steekproefresultaten ontwikkeld.

In de beginjaren van het commerciële marktonderzoek was de a-selecte steekproeftrekking de norm. Maar al vanaf de jaren 70 van de 20e eeuw bleek dat de a-selecte steekproef in veel gevallen niet of slechts tegen hoge kosten te realiseren was. Er ontstonden vormen van niet-kanssteekproeven, zoals de quotasteekproef. Bij de quotasteekproef worden respondenten geselecteerd op basis van kenmerken. In eerste instantie waren dit vooral demografische kenmerken, zoals geslacht, leeftijd, opleiding, functie en inkomen, waarvan populatiecijfers (censusdata) beschikbaar waren. De aanname was dat deze demografische kenmerken een sterke relatie hadden met het gedrag en de meningen van de doelgroep. Op deze wijze werd, zonder a-selecte trekking, een representatieve steekproef getrokken. Representativiteit (ofwel, vormt de steekproef een goede afspiegeling van de populatie waarover we iets willen zeggen) werd het criterium, en niet langer de a-selectie. Vanuit de theorie was dit zeer verdedigbaar en ook werd met behulp van vergelijkende studies aangetoond dat resultaten vergelijkbaar waren met 'echte' representatieve steekproeven, te meer daar de aard van de onderzoeken veranderde: minder beschrijvende doelstellingen, maar meer verklarend onderzoek en/of onderzoek naar verschillen tussen doelgroepen. Ofwel, minder vragen naar de totale consumptie van aardappelen in Nederland (daar zijn andere bronnen dan ondervraging geschikter voor), maar meer de vraag waarom mensen wel of niet aardappelen eten, op basis waarvan ze kiezen, etc..

Medio jaren 90 kwam online onderzoek op. Deze methode van dataverzameling ontwikkelde zich snel tot de meest toegepaste methode om kwantitatieve data te verzamelen. Dit kwam mede doordat in landen als Nederland de toegang tot het internet zich zeer snel ontwikkelde. Respondenten worden geworven uit eigen bestanden (bijvoorbeeld het klantenbestand) of uit grootschalige, speciaal hiervoor opgezette panels. Deze panels werven hun panelleden op diverse manieren, waarbij deels sprake was van zelfselectie: respondenten die zichzelf aanmelden. De steekproeftrekking uit de panels vindt veelal plaats met behulp van een quotabebanding. Respondentenpanels maken snel en goedkoop veldwerk mogelijk. Tegelijkertijd is er vanaf het begin discussie geweest over hoe representatief de mensen zijn die zich aanmelden voor panels. Ook het fenomeen 'professionele respondent' ontstond en zorgde voor veel discussie.

Representatief of relevant?

Als verklaring of verdediging van het gebruik van steekproeftrekking op basis van zelfselectie wordt vaak de argumentatie gebruikt dat het belangrijker is om relevante respondenten te ondervragen dan een representatieve steekproef te onderzoeken. Of zoals hieronder verwoord:

‘Pieter Willems, director PWNNext, noemt het de grootste verandering van de laatste twintig jaar en ingrijpender dan de transitie van telefonisch naar online onderzoek. ‘Het nieuwe onderzoeksparadigma is dat we van respondenten naar respons gaan, en van representativiteit naar relevantie. We samplen niet meer uit een universum maar uit relevantiepijlers: is een respons nuttig voor mijn onderzoek omdat het afkomstig is uit een gebied dat relevant is voor wat ik doe?’
(Clou 44, november 2009, ‘Van onderzoeken gaan we naar ondervinden’)

Dit lijkt te worden ondersteund door wat James Surowiecki schreef in het boek ‘*The Wisdom of Crowds*’, waarin hij aantoont dat de gemiddelde mening van groepen vaak meer accuraat de werkelijkheid weergeeft dan die van elk van de individuen in de groep. Hierbij moet wel bedacht worden dat ook hij uitgaat van groepen die minimaal enigszins representatief zijn voor de totale populatie. Relevantie is inderdaad een belangrijk criterium dat echter niet in plaats van, maar naast representativiteit geplaatst moet worden. Er kan alleen sprake zijn van relevantie indien de steekproef (aantoonbaar) een bepaalde mate van representativiteit heeft. Relevantie is een gevolg van de keuze van de doelgroep waaronder je onderzoek wilt doen; representativiteit is een gevolg van het steekproefkader en de manier waarop de steekproef daaruit wordt getrokken.

De discussie over relevant versus representatief is voor een groot deel ontstaan uit pragmatische overwegingen. Aan de ene kant is steeds meer informatie beschikbaar waarvan we niet precies weten hoe betrouwbaar de bron is en aan de andere kant is het steeds moeilijker om een (zo goed als) representatieve steekproef te trekken.

Naar onze mening kan het volgende gesteld worden:

- Voor elk onderzoek waarbij niet de gehele populatie onderzocht wordt, moet je de vraag stellen hoe belangrijk het is dat de steekproef lijkt op de populatie. Hoe belangrijk het is, verschilt naargelang de aard van de onderzoeksvraag. Bij puur exploratief onderzoek is het minder van belang, bij toetsend en beschrijvend onderzoek is het juist weer wel zeer belangrijk.
- Voor onderzoek met als doel het generaliseren van kwantitatieve uitkomsten naar de totale populatie, is de kanssteekproef theoretisch de beste optie.
- Kanssteekproeven zijn echter moeilijk te realiseren, met name door de non-respons. Niet-kanssteekproeven waarbij gewerkt wordt vanuit representativiteit vormen een goed alternatief voor de kanssteekproeven.
- Steekproeven waarbij gewerkt wordt met zelfselectie zijn goed bruikbaar als het minder belangrijk is dat de steekproef lijkt op de populatie of als met behulp van kenmerken van respondenten, liefst zowel demografische als psychografische, de steekproef op representativiteit wordt getoetst en aangepast.

2

Steekproeven: de hoofdzaak



Bij kwantitatief steekproefonderzoek willen we op basis van resultaten in de steekproef iets zeggen over een grotere populatie. Hoe kom je er achter welk percentage van de mensen in een bepaalde stad rookmelders geïnstalleerd heeft? Je kunt bij ieder huis aanbellen en noteren of er wel of geen rookmelder is. Een tijdrovende en kostbare aangelegenheid, zeker als het een grote stad is. Gelukkig hebben we hiervoor een oplossing: steekproeftrekking. De theorie leert ons dat als je van een bepaalde groep mensen een deel onderzoekt, de kans groot is dat wat je gevonden hebt onder die kleine groep, ook geldt voor de hele groep mensen. Hoe groot die kans precies is, kun je berekenen.

Vier begrippen zijn van belang als we het hebben over steekproeven. Ze worden nogal eens door elkaar gegooid, dus we beginnen met een definitie.

Betrouwbaarheid: De kans dat de uitkomsten van je onderzoek onder de steekproef hetzelfde zijn als wanneer je iedereen zou ondervragen. We rekenen meestal met een betrouwbaarheid van 95%. Dat wil dus zeggen dat als je 20 keer het onderzoek herhaalt, je 19 keer ongeveer dezelfde uitkomst zult krijgen.

Nauwkeurigheid: Dit heeft betrekking op het 'ongeveer' uit de vorige definitie. Als je in een onderzoek een bepaald percentage vindt, bijvoorbeeld 15%, dan weet je dat in werkelijkheid (onder de populatie) dit bijvoorbeeld tussen 13% en 17% ligt. Deze marge noemen we de nauwkeurigheidsmarge. Deze marge is afhankelijk van de betrouwbaarheid waarmee je iets wilt vaststellen. Hoe hoger de betrouwbaarheid, hoe groter de nauwkeurigheidsmarge. Ook is de nauwkeurigheid afhankelijk van de steekproefgrootte. Hoe groter de steekproef, hoe kleiner de nauwkeurigheidsmarge.

Significantie: Significantie heeft altijd betrekking op een verschil. Als de score van een bepaalde groep significant afwijkt van de score van een andere groep of van het totaal, betekent dit dat de kans zeer groot is dat als je het onderzoek onder de hele populatie zou uitvoeren in plaats van onder een steekproef, je ook een verschil zou vinden. Als je twee scores gevonden hebt die van elkaar afwijken en die afwijking is groot genoeg, dan kan er sprake zijn van een significant verschil.

Representativiteit: Een representatieve steekproef is een goede afspiegeling van de groep waarover je uiteindelijk uitspraken wilt doen, ofwel de populatie. Dit begrip staat los van de voorgaande drie begrippen, maar wordt er vaak mee verward.

Maurice de Hond of verkiezingen?

Je zou een verkiezing kunnen beschouwen als een groot onderzoek naar de politieke voorkeuren van een land. We hebben gezien dat de steekproeftheorie ons leert dat je niet iedereen hoeft te ondervragen om er toch achter te komen hoe iedereen ergens over denkt. Dit betekent dat als je een representatieve en grote steekproef zou trekken uit kiesgerechtigden en aan hen zou vragen waarop ze zouden stemmen, verkiezingen in principe overbodig zouden zijn (en dat zou een boel geld schelen).

In de praktijk weten we dat opiniepeilers er vaak naast zitten en ook nog eens onafhankelijk van elkaar tot andere conclusies komen. Hoe kan dat dan? Eén van de redenen die we verderop zullen behandelen, is dat de steekproef niet helemaal representatief is.

Wat is een 'goede' steekproef?

Voor we verdergaan, willen we eerst een misverstand uit de wereld helpen. Vaak wordt gedacht dat een grote steekproef automatisch een representatieve steekproef is, en dus een goede steekproef. Dit is geenszins het geval. Als een steekproef geen goede weergave is van de populatie waarover je iets wilt zeggen, maakt het niet uit hoe groot deze is, je zult nooit een goede steekproef hebben. Kijk eens naar de onderstaande voorbeelden en bedenk welke van deze steekproeven de beste is:

- 20.000 bezoekers van de website van het consumentenprogramma Kassa
- 500 mensen die tijdens het winkelen in de Kalverstraat op zaterdagmiddag worden aangesproken
- 25 willekeurig gekozen huisvrouwen die telefonisch benaderd worden

Het antwoord op de vraag wat de beste steekproef is, kunnen we zo niet geven. Dit wordt namelijk bepaald door de populatie waarover we uitspraken willen doen. Maar laat duidelijk zijn dat de website van Kassa vaak bezocht zal worden door zelfbewuste, kritische consumenten en daarmee vormen ze een afwijkende groep. Tenzij de populatie waarover je uitspraken wilt doen, de bezoekers van die website is. Hetzelfde geldt voor het winkelend publiek in de Kalverstraat. Wellicht enigszins representatief voor winkelend publiek in Amsterdam, maar niet representatief voor heel Nederland.

Het steekproefvraagstuk heeft een directe relatie met de keuze van de methode. De keuze van de methode wordt namelijk mede bepaald door de (on)mogelijkheid om een goede steekproef te trekken. Ook deze relatie zal in het volgende hoofdstuk aan de orde komen.

3

De representatieve steekproef



In kwantitatief onderzoek is het doel in de meeste gevallen inzicht verkrijgen in de mate waarin een bepaald verschijnsel voorkomt in een onderzoekspopulatie. Deze populatie kan groot zijn, bijvoorbeeld alle Nederlanders, of klein, bijvoorbeeld alle bezitters van een Ferrari in Zuid-Holland. In de meeste gevallen is het niet mogelijk of wenselijk de hele populatie te onderzoeken, dus wordt een steekproef onderzocht.

Via onderzoek willen we bijvoorbeeld vaststellen welk percentage van alle Nederlandse mannen van 18 jaar of ouder een bril draagt. Hiertoe ondervragen we een steekproef die getrokken is uit deze doelgroep. Stel dat we in de steekproef 24% bril dragers aantreffen, dan concluderen we dat van alle Nederlandse mannen van 18 jaar en ouder (naar schatting) 24% een bril draagt. Merk op dat gesproken wordt over een schatting. We komen hierop terug.

Vanuit resultaten van onderzoek onder een steekproef conclusies trekken over een populatie mag alleen als de steekproef representatief is. Als niet aangetoond of aannemelijk gemaakt kan worden dat de steekproef representatief is voor de populatie, mogen we wel vaststellen dat 24% van de steekproef een bril draagt, maar niet daaruit concluderen dat ook 24% van de populatie een bril draagt.

Representatief wil zeggen dat de steekproef een goede afspiegeling is van de populatie op het kenmerk dat we willen meten. Dit is curieus: We willen een kenmerk meten en stellen als eis dat de steekproef er hetzelfde uitziet als de populatie (op de kenmerken die we niet kennen). Hoe lossen we dit op? Hoe komen we dan aan een representatieve steekproef? Er zijn twee manieren van aanpak om te komen tot een steekproef die een hoge kans heeft om representatief te zijn:

- De a-selectie benadering
- Representatief op kenmerken die wel bekend zijn

De a-selectie benadering baseert zich op de wetten uit de waarschijnlijkheidsleer, en wel specifiek de wet van de grote getallen. Deze wet stelt dat als uit een populatie op een volledig willekeurige wijze een steekproef wordt getrokken en de steekproef 'enige omvang' heeft, het resultaat van de steekproef waarschijnlijk dicht in de buurt zit van de werkelijkheid binnen de populatie. Wat 'enige omvang' is, is niet exact aan te geven. In de praktijk van het marktonderzoek wordt echter uitgegaan van een minimale omvang van 60 tot 100 respondenten. Hoe nauwkeurig het resultaat is, wordt aangegeven met de begrippen (statistische) betrouwbaarheid en nauwkeurigheid. De statistiek geeft ons middelen om te berekenen hoe nauwkeurig de schatting is bij een gewenste betrouwbaarheid. Hier geldt: hoe groter de (representatieve) steekproef, hoe nauwkeuriger de schatting.

Bij de tweede aanpak wordt uitgegaan van kenmerken van de doelgroep die wel bekend zijn en zorgen we ervoor dat de steekproef op deze kenmerken dezelfde samenstelling heeft als de populatie. Als onderzoek uitgevoerd wordt onder 'gemiddeld' Nederland, zijn gegevens bekend over de verdeling binnen de populatie naar geslacht, leeftijd, regio, opleiding en andere kenmerken. Er wordt vervolgens aangenomen dat, als ervoor gezorgd wordt dat de verdeling op deze kenmerken in de steekproef gelijk is aan die binnen de totale populatie, de steekproef ook representatief is op de kenmerken die we willen onderzoeken. Als er gewerkt wordt met voldoende bekende kenmerken, is dit een redelijke aanname, maar niet helemaal waterdicht. Het representatief verklaren van een steekproef met behulp van bekende kenmerken mag alleen als meerdere kenmerken gebruikt worden en de steekproef wordt getrokken uit een neutrale en brede bron.

Achteraf, dus nadat het onderzoek is uitgevoerd en de steekproef is gerealiseerd, kan een controle worden uitgevoerd op de representativiteit ervan. Dit doen we door kenmerken waarvan we weten hoe vaak die voorkomen binnen de populatie te leggen naast kenmerken waarover we de steekproef hebben ondervraagd. In ieder geval moeten we hier dus kenmerken in de vragenlijst opnemen waarvan populatiecijfers bekend zijn. Dit kunnen demografische kenmerken zijn, maar ook gedragskenmerken. Als er sprake is van wezenlijke verschillen op kenmerken tussen de gerealiseerde (netto) steekproef en de populatie, is herwegen een optie.

Hoe kom je aan populatiecijfers?

Om achteraf te controleren of je steekproef representatief is of om vooraf te bepalen hoe die er uit zou moeten zien, heb je informatie nodig over hoe de populatie er uitziet. Daarvoor kun je een aantal bronnen gebruiken:

- CBS: De website van het CBS (www.statline.nl) kan je veel leren; vooral als het gaat om informatie over geheel Nederland kun je hier vaak terecht.
- De Gouden Standaard: Veel marktonderzoeksbureaus en opdrachtgevers die lid zijn van de MOA maken gebruik van deze ijkingsmethode. De Gouden Standaard bevat populatiecijfers over allerlei soorten achtergrondkenmerken van Nederlanders. Denk aan werksituatie, huishouden, opleiding etc., op landelijk, maar ook op gemeentelijk niveau.
- Eerdere onderzoeken en interne data: Als een bepaalde opdrachtgever vaker onderzoek gedaan heeft en/of interne informatie beschikbaar heeft, dan is dit vaak een goed startpunt. Bedenk wel dat als je uitkomsten van een ander onderzoek als ijkpunt neemt, je goed moet weten in hoeverre dat onderzoek gedaan is onder een representatieve steekproef.

Soms is er helemaal niets bekend over de populatie. Dan zal je moeten vertrouwen op het a-select benaderen van respondenten en/of moeten accepteren dat je achteraf niets kunt zeggen over de mate van representativiteit van je steekproef.

4

De stappen van steekproeftrekking



Voor het trekken van een steekproef worden de volgende stappen gezet:

1. Definieer de doelgroep voor onderzoek
2. Bepaal het steekproefkader: de bron waaruit de steekproef wordt getrokken
3. Bepaal de manier van steekproeftrekking
4. Bepaal de gewenste steekproefgrootte
5. Bepaal of en hoe je moet herwegen

1. Definieer de doelgroep en opbouw van de steekproef

Als het goed is, heb je dit al gedaan in de fase van de probleemanalyse. Daar omschrijf je de doelgroep zo nauwkeurig mogelijk aan de hand van beschrijvende kenmerken, en waar mogelijk in aantallen. Handig is ook om alvast na te denken over subgroepen, die je wilt bekijken in de analysefase. Je denkt dus vooruit en je stelt hypothesen op. Tussen welke doelgroepen verwacht je verschillen te vinden? Denk bijvoorbeeld aan gebruikersgroepen: niet-gebruikers van een product, incidentele gebruikers en regelmatige gebruikers. De definitie van de doelgroep vormt ook de basis van de vragen die je stelt om te controleren of een respondent voldoet aan de criteria om mee te doen aan je onderzoek, de selectievragen.

Je kunt een hele lijst met kenmerken opstellen op basis waarvan je wilt dat de steekproef representatief is. In de praktijk is zo'n fijnmazige steekproef niet haalbaar. Bedenk dat je als de steekproefcellen 'interlocked' ofwel op elkaar inhakend definieert, het aantal cellen exponentieel toeneemt. Dus als je een steekproef wilt die representatief is op geslacht, leeftijd in 4 categorieën, inkomen in 3 categorieën en 5 regio's, heb je $2 \times 4 \times 3 \times 5 = 120$ cellen die je moet vullen. Bedenk daarom vooraf goed welke criteria je als het meest relevant beschouwt als het gaat om de representativiteit van de steekproef.

Een aandachtspunt bij steekproeven waarbij de te onderzoeken eenheden uit meerdere personen bestaat, is de vraag wie je precies wilt ondervragen. Wil je binnen huishoudens degene die de meeste boodschappen doet? Degene die (mede) verantwoordelijk is voor financiële beslissingen? Wil je binnen een bedrijf het hoofd HR, of de verantwoordelijke voor het wagenpark? Etc..

In deze stap kijk je ook vooruit naar stap 5, waarin je gaat kijken of een herweging nodig is. Daarvoor is het namelijk nodig dat je weet hoe de populatie er in werkelijkheid uitziet. Je moet dus op zoek naar populatiegegevens. Het is handig om dit vooraf te doen, omdat je dan je steekproef exact kunt definiëren op basis van bekende criteria. Doe je dit niet, dan loop je de kans dat je er achter komt dat je geen populatiecijfers beschikbaar hebt, of dat die op een andere manier zijn gedefinieerd (bijvoorbeeld 5 opleidingsniveaus, terwijl je in je vragenlijst vraagt naar 3 niveaus).

2. Bepaal het steekproefkader

Het steekproefkader is de bron waaruit de steekproef getrokken wordt. In deze stap denk je dus na over waar je je doelgroep vindt. De bron van je bruto steekproef kan zijn een 'lijst' of een plek waar de doelgroep samenkomt. Voorbeelden van lijsten: een klantenbestand, het telefoonboek, de Gemeentelijke Basis Administratie, een consumentenpanel, het bestand van een selectiebureau. Voorbeelden van plekken: op straat, in winkelcentra, op evenementen, op websites. Een goed steekproefkader is volledig, is representatief voor de doelgroep of het is mogelijk er een representatieve steekproef uit te trekken

Bedrijven die direct zaken doen met hun eindafnemers beschikken vaak over een klantenbestand. Als de doelgroep voor onderzoek bestaat uit klanten, bijvoorbeeld bij klanttevredenheidsonderzoek, is dit een goed steekproefkader. Dit geldt alleen als kan worden vastgesteld dat het bestand volledig is en/of je de beschikking krijgt over het gehele bestand om daaruit een steekproef te trekken. Soms willen opdrachtgevers zelf eerst een selectie maken van klanten die benaderd kunnen worden. Een opdrachtgever met 20.000 klanten kan het bezwaarlijk vinden het hele bestand voor onderzoek ter beschikking te stellen. In dat geval kan hij een steekproef van 1.000 adressen uit het volledige bestand trekken, waaruit de uiteindelijke steekproef getrokken kan worden. Dit is prima, zolang dit a-select gebeurt. Als de opdrachtgever klanten selecteert waarvan hij weet dat die graag mee willen werken, is de kans groot dat hij de meest positieve klanten selecteert en daarmee de uitkomsten ernstig vertekent. Dit verschijnsel doet zich vaker voor dan je denkt, vooral als de selectie wordt gedaan door personen die belang hebben bij positieve onderzoeksuitkomsten (bijvoorbeeld accountmanagers).

Bij het gebruik van (adressen)bestanden moet dus in eerste instantie bekeken worden of deze wel de volledige doelgroep afdekken. Iedereen zal kunnen begrijpen dat een steekproef die getrokken wordt uit lezers van het Reformatorisch Dagblad een ander onderzoeksresultaat zal opleveren bij een onderzoek naar politieke voorkeur, dan eenzelfde soort steekproef uit lezers van het dagblad Trouw, ook al is de verdeling van de kenmerken geslacht, leeftijd en regio gelijk aan de verdeling van alle Nederlanders in beide steekproeven. In geen van beide situaties is het steekproefkader representatief voor de Nederlandse kiezer. Resultaten kunnen alleen geldig zijn voor de doelgroepen 'lezers van'.

De keuze van de methode van onderzoek wordt in het geval van beschikbaarheid van lijsten mede bepaald door de beschikbaarheid van soorten adresgegevens in beschikbare lijsten: adres, telefoonnummer, e-mailadressen. Hierbij kan sprake zijn van onvolledigheid van bepaalde gegevens. Bijvoorbeeld, het bestand bevat slechts 40% mensen met een e-mailadres. Je moet je dan afvragen of het een specifieke (niet-representatieve) groep is waarvan de e-mailadressen beschikbaar zijn. Als in dat geval wel van alle klanten een telefoonnummer beschikbaar is, zal overwogen moeten worden om de klanten (eerst) telefonisch te benaderen.

Bij online onderzoek wordt vaak gebruikgemaakt van grootschalige consumentenpanels. Grotere onderzoeksbureaus hebben eigen panels en er zijn gespecialiseerde panelbureaus die een panel beheren. In de panels zitten mensen die hebben aangegeven bereid te zijn mee te werken aan onderzoek. Van deze personen zijn vaak diverse gegevens bekend, zodat selectie van specifieke doelgroepen mogelijk is. Het steekproefkader, in dit geval het panelbestand, vormt hier een belangrijk aandachtspunt. Is het panelbestand representatief voor de te onderzoeken populatie, of is een representatieve steekproef uit dit bestand te trekken? Panelbeheerders houden meestal al rekening met representativiteit bij de selectie van hun panelleden, controleren periodiek de representativiteit van hun bestand en houden bij het benaderen van hun panelleden al rekening met de representativiteit. Om de kwaliteit van online panels te beoordelen, heeft Esomar een lijst opgesteld met 28 vragen die een indicatie geven van de manier van werken en de representativiteit van het panel. Bij de meeste Nederlandse en internationale panels kun je de antwoorden op deze [28 vragen](#) opvragen.

Als geen goed bestand beschikbaar is, kan een eigen steekproefkader gecreëerd worden door mensen of hun contactgegevens te werven op specifieke plekken. Vliegreizigers worden geworven op vliegvelden, winkelbezoekers bij de uitgang van een winkel en bezoekers van evenementen bij het verlaten van het

evenement. Rijkswaterstaat noteert voor onderzoek onder gebruikers van bepaalde wegen kentekens en zoekt er de adresgegevens bij. Hierbij is het zaak om enerzijds te bekijken of het steekproefkader volledig of representatief is en anderzijds de werving ter plekke op een a-selecte manier uit te voeren. Als de doelgroep voor onderzoek bestaat uit de bezoekers van een (eenmalig) evenement, kan een volledig steekproefkader gecreëerd worden door bij de uitgang te staan gedurende de gehele periode dat mensen het evenement kunnen verlaten. Bij werving op openbare plekken zal vaak geen volledig steekproefkader gecreëerd kunnen worden. Er moet dan wel gezorgd worden voor representatieve plekken (niet alleen werving in de Kalverstraat in Amsterdam) en voor een goede spreiding van momenten (overdag, 's avonds, doordeweeks en in het weekend). Zelfs als de plek en de momenten als representatief kunnen worden aangemerkt, zit hier een valkuil, die beschreven wordt in onderstaand voorbeeld.

Een onderzoeksbureau ging een onderzoek uitvoeren onder vliegreizigers. De doelgroep voor onderzoek was omschreven als 'mensen die het afgelopen jaar minimaal één vliegreis hebben gemaakt'. Uit deze doelgroep wilde men een representatieve steekproef ondervragen. Er was geen bestand beschikbaar van deze doelgroep, dus werden respondenten geworven op een groot vliegveld. De steekproeftrekking gebeurde keurig op een volledig a-selecte wijze op het vliegveld. De bereidheid om mee te werken was groot en zo ondervroeg men 500 vliegreizigers.

Hoe representatief is de steekproef van 500 reizigers? Bij een eerste snelle blik op de resultaten bleek 33% van de ondervraagden frequent vliegreiziger te zijn (minimaal 5 vliegreizen per jaar). Uit landelijk onderzoek was echter bekend dat slechts 2% van alle Nederlanders 5 of meer vliegreizen per jaar maakt. De steekproef was dus blijkbare niet representatief voor de doelgroep, terwijl deze toch zo netjes getrokken was. Wat is hier aan de hand? De oorzaak van dit resultaat is dat frequente vliegreizigers nu eenmaal vaker op een vliegveld komen dan niet-frequente reizigers. Daarmee hebben ze een grotere kans om in de steekproef terecht te komen. De steekproeftrekking voldoet dus niet aan de eis dat iedereen in de populatie een even grote kans heeft om in de steekproef terecht te komen. We kunnen nu drie kanten op:

- **Herdefinitie van de doelgroep:** mensen op een vliegveld. We accepteren dan dat frequente reizigers een grotere stem hebben in de resultaten (wat geen gekke gedachte is als je onderzoek bijvoorbeeld over winkels en Horeca op het vliegveld gaat).
- **Herwegen van de resultaten:** frequente reizigers krijgen een lage weegfactor en niet-frequente reizigers een hoge. Dit kunnen we doen als we cijfers over het vlieggedrag van de Nederlander hebben.
- Als we dit vooraf aan hadden zien komen, dan hadden we met behulp van het **instellen van quota** kunnen voorkomen dat de steekproef scheef zou worden. We zouden dan categorieën kunnen maken op basis van de frequentie waarmee men vliegt. De interviewers zouden dan een turflijstje meekrijgen waarop staat hoeveel mensen uit elke categorie ze zouden moeten interviewen.

De beperking van het werven ter plekke zit dus in de mogelijkheid dat bepaalde subgroepen vaker 'ter plekke' zijn dan andere en daarmee een hogere kans hebben om in de steekproef terecht te komen. Als deze mensen afwijken op een relevant kenmerk, hebben we dus een probleem. Als we bijvoorbeeld een onderzoek doen onder werkende vrouwen, missen we de doelgroep die voltijds werkt als we doordeweeks op straat gaan interviewen.

Je kunt in deze stap er achter komen dat er geen geschikt steekproefkader te vinden is om een goede steekproef uit te trekken. Misschien kun je de mensen niet vinden, niet bereiken, het wordt veel te duur of zijn het er gewoon niet genoeg om een kwantitatief onderzoek uit te voeren. In zo'n geval moet je terug naar de tekentafel en je keuze voor de onderzoeksmethodologie heroverwegen. Wellicht kunnen de onderzoeksvragen ook door middel van kwalitatief onderzoek onder een kleine steekproef worden beantwoord? Wellicht bieden expertinterviews uitkomst.

We kregen een onderzoeksaanvraag over begraafplaatsen. Doelstelling was het in kaart brengen van het keuzeproces als het gaat om de keuze voor een bepaalde begraafplaats. Belangrijkste onderzoeksvraag was: Welke overwegingen bij de keuze van een begraafplaats zijn belangrijk voor de nabestaanden?

Omdat het om een keuzeproces ging, zou de ideale opzet zijn een onderzoek onder mensen die net dit keuzeproces hadden doorlopen. Deze mensen bereid vinden om mee te doen aan iets triviaals als een marktonderzoek vlak na het overlijden van een naaste, zou echter problematisch zijn (en bovendien onwenselijk uit ethisch oogpunt). Als alternatief hebben we een onderzoek voorgesteld onder begrafenisondernemers en -begeleiders. Deze experts hebben dagelijks te maken met keuzeprocessen en spelen daarin bovendien een belangrijke adviserende of zelfs sturende rol.

3. Bepaal de manier van steekproeftrekking

Voor de manier van steekproeftrekking wordt een onderscheid gemaakt tussen kanssteekproeven en niet-kanssteekproeven.

Kanssteekproeven (ofwel a-selecte steekproeftrekking)

- Volledig a-selecte steekproef: Het steekproefkader bestaat uit alle elementen uit de populatie en daaruit wordt op een volledig a-selecte wijze een steekproef getrokken, zoals dat gebeurt bij de trekking van een loterij. Elk element heeft een even grote kans om in de steekproef terecht te komen.
- Systematische steekproef: Ook hier zijn alle elementen uit de populatie bekend, maar in plaats van volledig willekeurig (a-select) te trekken, wordt elk zoveelste element uit de bekende lijst getrokken. Dit wordt bijvoorbeeld vaak toegepast als van de populatie een vaste lijst bestaat, zoals een klantenbestand. Er is bijvoorbeeld een lijst van alle klanten in alfabetische volgorde. Als er 20.000 klanten op de lijst staan en we hebben 1.000 respondenten nodig als bruto steekproef, dan selecteren we elke 20^{ste} klant op de lijst. Als aangenomen mag worden dat de volgorde in de lijst geen relatie heeft met de te onderzoeken variabelen (wat bij een alfabetische volgorde een realistische aanname is), kan een systematische steekproef beschouwd worden als een volledig a-selecte steekproef. Zou je echter de eerste 1.000 van de lijst nemen, dan loop je kans op een vertekening. Dit kan bijvoorbeeld gebeuren doordat van oorsprong Nederlandse namen vaak met 'de' of 'van' beginnen en dus relatief vaak vooraan en achteraan het alfabet zijn te vinden.

Varianten hierop zijn:

- De getrapte steekproef: wordt in verschillende stappen getrokken, nog steeds vanuit de a-selecte benadering. Bijvoorbeeld eerst een a-selecte steekproef uit alle gemeenten, gevolgd door een a-selecte steekproef van straten, en binnen die straten een a-selecte steekproef van huisadressen.
- De gestratificeerde steekproef: hierbij wordt de populatie op voorhand ingedeeld op kenmerken, waarbij uit elke subgroep een steekproef wordt getrokken. Dit kan naar rato van het aantal elementen binnen elke subgroep (proportioneel) of anders (disproportioneel). Er zijn twee redenen voor deze wijze van steekproeftrekking: er zijn verschillende bronnen, of we willen er zeker van zijn dat bepaalde subgroepen in voldoende mate in de steekproef aanwezig zijn. In dat laatste geval zal sprake zijn van een disproportioneel gestratificeerde steekproef. Bij onderzoek waarbij de doelgroep bestaat uit bedrijven zal hiervoor vaak gekozen worden. Een volledig a-selecte steekproef van bedrijven zal slechts een zeer beperkt aantal grote bedrijven opleveren, omdat er nu eenmaal veel meer kleine bedrijven zijn dan grote. Dit terwijl we vaak juist geïnteresseerd zijn in de mening van grote bedrijven. De totaalresultaten zullen wel moeten worden herwogen.

Niet-kanssteekproeven

- De meest voorkomende variant hiervan is de quotasteekproef. Hierbij worden niet willekeurig mensen uit de populatie uitgenodigd voor het onderzoek, maar selecteert de onderzoeker of interviewer mensen die aan bepaalde kenmerken voldoen, en wel op zo'n manier dat de totale steekproef een gewenste samenstelling krijgt op de bekende kenmerken. Hiermee wordt representativiteit nagestreefd.

Andere manieren van niet-kanssteekproeven zijn:

- Gemakssteekproef (convenience sampling): de selectie van respondenten vindt plaats op een manier die het meest eenvoudig is voor de onderzoeker.
- Judgement steekproef: de onderzoeker selecteert respondenten op basis van geschiktheid voor het onderzoek.
- Zelfselectie: respondenten melden zichzelf aan.
- Snowball sample: de onderzoeker begint met een aantal individuen uit de doelgroep en vraagt vervolgens aan hen of ze de onderzoeker in contact kunnen brengen met andere personen uit de doelgroep.

Het zal duidelijk zijn dat bij deze wijzen van steekproeftrekken de representativiteit van de steekproef in bijna alle gevallen discutabel is. Toch wordt uit pragmatische overwegingen regelmatig gekozen voor één van deze manieren van steekproeftrekken. Zolang rekening wordt gehouden met de beperkte representativiteit en de eventuele gevolgen daarvan, hoeft dit geen groot bezwaar te zijn.

4. Bepaal de gewenste steekproefgrootte

Als voorlaatste stap ga je bepalen hoe groot de steekproef zou moeten zijn. De gewenste netto steekproefgrootte wordt bepaald door de afweging van haalbaarheid en kosten versus de gewenste betrouwbaarheid en door de wens om uitspraken te kunnen doen over bepaalde subgroepen. Hoe groter je steekproef, hoe nauwkeuriger je resultaten, maar ook hoe duurder je onderzoek.

De belangrijkste factoren die de steekproefomvang bepalen, zijn¹:

- a. Hoe nauwkeurig wil je uitspraken kunnen doen?
- b. Over hoeveel subgroepen wil je uitspraken kunnen doen?

a. Hoe nauwkeurig wil je uitspraken kunnen doen?

Als steekproefresultaten (achteraf) op de populatie geprojecteerd worden, zijn dit altijd schattingen. Vergelijk het met een finishfoto die wordt gebruikt om te bepalen wie de winnaar van een wielervedstrijd. Een scherpe foto toont de winnaar. De foto die een marktonderzoek oplevert, is echter altijd enigszins onscherp. Hoe groter de steekproef, hoe scherper de foto en dus hoe beter we kunnen vaststellen of er daadwerkelijk een verschil is. De statistiek geeft ons middelen om te bepalen hoe 'goed' de schatting is. De kwaliteit van de schatting wordt bepaald met de begrippen betrouwbaarheid (hoe groot is de kans dat we een juiste uitspraak doen?) en nauwkeurigheid (wat is de marge rondom de gevonden steekproefuitkomst?).

¹ Er zijn theoretisch nog meer factoren, namelijk de gewenste betrouwbaarheid van de uitspraak en de verwachte uitkomst. In de praktijk nemen we voor de betrouwbaarheid 95% en voor de verwachte uitkomst 50%, (de steekproefcalculator doet dit ook), maar dit hoeft niet altijd. De benodigde steekproef om met een bepaalde nauwkeurigheidsmarge een uitspraak te doen is bij een verwachte uitkomst van 10% veel kleiner dan bij 50%. Dus stel dat je 1 doel hebt, namelijk testen of minimaal 10% van de populatie iets wil, denkt of doet, dan kun je met een kleinere steekproef volstaan.

Gesteld dat je in een onderzoek met een representatieve steekproef van $n=100$ vindt dat 25% van de automobilisten in je steekproef van plan is winterbanden aan te schaffen, wat kun je dan zeggen over de populatie van automobilisten? Je kunt zeggen, met een betrouwbaarheid van 95%, dat naar schatting tussen de 17% en 33% van de Nederlandse automobilisten van plan is winterbanden aan te schaffen. De onnauwkeurigheidsmarge is 8% (afgerond). Vind je deze marge te groot, dan kun je de steekproef verhogen. Is de steekproef bijvoorbeeld $n=500$, dan is de marge nog maar 4% (de schatting is dan tussen de 21% en 29%).

Bij een andere benadering kijk je naar welke verschillen je als significant wilt kunnen bestempelen. Deze benadering is bijvoorbeeld van toepassing als je een 0- en een 1-meting wilt doen. Je kunt uitrekenen wat de steekproefgrootte zou moeten zijn als je bijvoorbeeld een stijging van 5% in de bekendheid als significant wilt kunnen bestempelen. Om deze berekening te kunnen maken, verwijzen we naar de statistische literatuur.

Op de website www.allesovermarktonderzoek.nl vind je een steekproefcalculator die je vertelt welke steekproefomvang je nodig hebt om een bepaalde nauwkeurigheid te bereiken. Deze steekproefcalculator biedt de mogelijkheid om de grootte van de populatie mee te nemen in je berekening.

b. Over hoeveel subgroepen wil je uitspraken kunnen doen?

In theorie geldt het bovenstaande voor elke subgroep waarover uitspraken gedaan moeten worden. Voor elke subgroep is dus een gewenste steekproefomvang te berekenen. In de praktijk zal dit veelal niet gebeuren, maar wordt gewerkt met een minimale steekproef per subgroep. In de praktijk rekenen marktonderzoekers vaak met een minimale omvang van tussen de 60 en 100 respondenten. Om je totale steekproef te bepalen, zou je het minimum per subgroep kunnen vermenigvuldigen met het aantal subgroepen waarover je iets wilt kunnen zeggen. Let wel op dat je de uiteindelijke steekproef wel moet herwegen om tot een goede totale steekproef te komen.

5. Bepaal of en hoe je moet herwegen

Feitelijk hoort herwegen bij de analysefase, maar omdat het alles te maken heeft met je steekproef, behandelen we het hier. Als je een steekproef hebt getrokken die afwijkt van de populatie, kun je dit (deels) recht trekken door herweging. Bij herwegen tel je bepaalde respondenten zwaarder of juist minder zwaar mee in je berekeningen. Je gebruikt herweging ook om bij een quotasteekproef totaalscores voor de hele populatie uit te kunnen rekenen.

Hoewel herwegen iets is wat je achteraf doet, zal je er vooraf al rekening mee moeten houden dat het eventueel nodig is. Je zult namelijk precies de verhoudingen in de populatie moeten weten om überhaupt te kunnen herwegen.

Bedenk dat herwegen geen heilige graal is, die ervoor zorgt dat je elke steekproef achteraf representatief kunt maken. Het is eerder een lapmiddel waarmee je een scheve steekproef enigszins kunt herstellen. Je moet met name oppassen als je doelgroepen waarvan je relatief weinig mensen ondervraagd met een hoge weegfactor zwaarder laat meetellen. Met het herwegen, vergroot je namelijk ook de ruis en vertekening die een kleine steekproef potentieel in zich heeft. Dus als je wilt berekenen of er een significant verschil is tussen 2 subgroepen uit de totale steekproef moet dit op basis van de ongewogen data.

5

Het issue non-respons



Om de gewenste netto steekproefgrootte te realiseren, zal een grotere bruto steekproef uitgezet moeten worden. Het verschil tussen de bruto en de netto steekproef is de non-respons. Non-respons ontstaat doordat mensen die benaderd worden om mee te werken aan onderzoek niet kunnen of willen meewerken. Het probleem hierbij is dat er een kans bestaat dat er sprake is van selectieve non-respons, dat wil zeggen dat degenen die niet meewerken andere resultaten zouden geven dan degenen die wel meewerken.

Stel dat we onderzoek willen doen naar de bereidheid van mensen om onbekenden te helpen die op straat de weg vragen. We willen 200 mensen ondervragen en we doen dit door middel van straatinterviews. Om aan 200 deelnemers te komen, moeten we 400 mensen aanspreken. 50% Wil dus niet meewerken aan het interview. De kans is groot dat die 50% die niet wil meewerken aan het interview minder geneigd is om onbekenden op straat te helpen. Dit noemen we selectieve respons.

Hierdoor is het steekproefresultaat niet meer representatief voor de populatie en mogen we dus eigenlijk niet meer de steekproefresultaten direct hierop projecteren. Als aangetoond kan worden dat de non-respons volledig a-select is, hoeft dit echter geen probleem te zijn. Het aantonen hiervan zal echter jammer genoeg niet of nauwelijks mogelijk zijn.

Niet mee willen werken, ofwel weigering, speelt bij alle methoden van ondervraging. Over de volle breedte is de bereidheid om mee te werken aan marktonderzoek gedaald ten opzichte van enige jaren geleden. Dat is een serieuze bedreiging voor de representativiteit van steekproeven en daarmee voor de kwaliteit van onderzoek. De bereidheid om mee te werken aan onderzoek hangt af van de doelgroep, het onderwerp, de methode van onderzoek en de manier van benaderen. Bij een juiste aanpak zijn in veel gevallen nog steeds goede responspercentages te realiseren. Dit vraagt echter wel tijd, aandacht en soms geld. Bij moeilijke steekproeven, waarbij een hoge respons belangrijk is, is het zinvol te overwegen om meerdere mogelijkheden (methoden) om mee te werken aan te bieden.

Voor een grootschalig onderzoek onder de bevolking stuurde het CBS brieven naar een a-selecte steekproef van Nederlanders. In de brief werd (na een uitgebreide toelichting) gevraagd om mee te werken. In eerste instantie via online onderzoek. Via een link in de brief kon men bij de vragenlijst op internet komen. Daarnaast werd de mogelijkheid geboden om een schriftelijke vragenlijst aan te vragen (antwoordvelop was ingesloten). Tenslotte werd de mogelijkheid geboden voor een persoonlijk interview. Hiervoor werden de benaderde personen die niet gereageerd hadden allemaal telefonisch benaderd. De telefonische benadering was enerzijds bedoeld om nog personen over te halen de online of schriftelijke vragenlijst in te vullen. Lukte dit niet, dan werd

geprobeerd het interview telefonisch af te nemen (beperkte vraagstelling) of een afspraak te maken voor een mondeling interview. Op deze wijze werd er alles aan gedaan om een hoge respons te realiseren.

Een onderzoek dat leuk is voor respondenten en waar een redelijke beloning tegenover staat, zal een hogere respons opleveren dan een onderzoek dat langdradig en saai is. Een respondent die ooit meegedaan heeft aan een leuk onderzoek, zal de volgende keer dat hij of zij benaderd wordt eerder weer meedoen. Gamification, ofwel de ontwikkeling dat respondenten onderzoeken ervaren alsof ze meedoen aan een spel, kan zorgen voor hogere deelname bereidheid aan onderzoek.

Het probleem met zelfselectie

Het probleem van een niet-representatieve steekproef is heel duidelijk in het geval van zelfselectie van respondenten. Deze melden zich hierbij zelf aan om mee te werken aan een onderzoek. De kans is hierbij groot dat degenen die zich spontaan aanmelden niet representatief zijn voor de totale populatie. Onderzoek heeft aangetoond dat bij zelfselectie meningen meer gespreid zijn: met name de aperte voor- en tegenstanders melden zich aan. Dit kan gezien worden als een responsprobleem, maar het gaat iets breder. Het is veelal een combinatie van steekproefkader en respons. Vreemd genoeg zijn het vaak onderzoeken met een politiek motief die aan het euvel van zelfselectie leiden. MOA-ombudsman Lex Olivier tikte in 2010 en 2011 onder meer de SP en de NRC op de vingers voor onderzoeken waarin sprake was van een selectieve steekproef. Een ander voorbeeld is het ANWB-onderzoek uit 2009 over het maatschappelijk draagvlak van rekeningrijden. Hierover schreven we eerder op onze blog het volgende:

Draagvlak voor draagvlakmeting ontbreekt

Hulde voor minister Eurlings die maatschappelijk draagvlak als voorwaarde stelt voor een nieuwe wet die nogal wat gevolgen lijkt te gaan hebben. Maar hoe meet je maatschappelijk draagvlak? Dit laten doen door een belangenorganisatie en je eigen politieke toekomst hiervan afhankelijk te maken, is onverstandig. De ANWB lijkt wel een verstandige keuze te maken door niet te gaan voor een populistische JA/NEE peiling, maar naar een onderbouwd oordeel over de mogelijke gevolgen van de invoering van de wet te vragen. Toch gaat er bij de ANWB-peiling nogal wat mis, waardoor de uitkomsten niet bruikbaar zijn in maatschappelijk en politiek debat. De ANWB-peiling zal dit eerder verstoren dan er een zinvolle bijdrage aan leveren.

De steekproef

Het belangrijkste probleem is niet de gekozen vraagstelling, maar de manier van steekproef trekken. Inmiddels hebben honderdduizenden respondenten de vragenlijst ingevuld. Een grote steekproef, dus representatief, wordt algemeen geredeneerd. De media houden van grote steekproeven. Niet correct dus. Van representativiteit kan geen sprake zijn bij een dergelijke opzet waarbij iedereen die dat wil de vragenlijst kan invullen. Sterker nog, degenen met de meest uitgesproken meningen zullen eerder geneigd zijn de moeite te nemen de vragenlijst in te vullen. Meestal zijn het de tegenstanders die het hardst roepen, dus van een representatieve steekproef kan niet gesproken worden. Doordat vooraf is aangekondigd welk belang gehecht wordt aan de uitkomsten van de peiling, wordt dit effect alleen maar versterkt. Voornamelijk is onduidelijk of de steekproef op de een of andere manier wordt vergeleken met een representatief getrokken steekproef.

De ANWB beschikt over een groot ledenbestand. Een mooie, zorgvuldige steekproef getrokken uit dit bestand zou een veel representatiever beeld geven van de mening van de Nederlandse automobilist. Uit het oogpunt van betrouwbaarheid heeft het weinig zin een steekproef groter dan $n=2.000$ te trekken. Het levert nauwelijks een betrouwbaarder beeld op, zolang de steekproef maar zorgvuldig getrokken is. Helaas doet een 'steekproef' van enkele honderdduizenden het beter in de media.

Respondenten die niet hun best doen

Een specifieke vorm van non-respons zijn frauduleuze respondenten. Wellicht een groter probleem dan respondenten die niet aan een onderzoek mee willen doen, zijn de respondenten die wel meedoen, maar dit niet serieus doen. De komst van onderzoekspanels heeft ertoe geleid dat het gebruikelijk is dat respondenten een klein geldbedrag krijgen voor deelname aan onderzoek. En dit leidt er weer toe dat sommige mensen vooral meedoen vanwege het geld. Dat is op zichzelf niet erg, zolang ze maar serieus meedoen. En dat is lang niet altijd het geval. Uit ervaring blijkt dat normaal gesproken tussen 5% en 10% van de deelnemers aan een online onderzoek dit niet serieus doen. Als een interview wordt uitgevoerd door een interviewer die vragen stelt, dan kan die interviewer redelijk inschatten of een respondent serieus is en eventueel de respondent motiveren de vragenlijst serieus af te maken. Als het onderzoek wordt uitgevoerd via internet, wordt het lastig om tijdens het interview te controleren of de respondent serieus is.

Om te zien of iemand zijn of haar best heeft gedaan bij het invullen van de vragenlijst, zijn er drie belangrijke indicatoren:

1. Hoe snel vult de respondent de vragenlijst in?
2. Hoe consequent is de respondent in zijn antwoorden?
3. Hoe gemakzuchtig is de respondent bij het beantwoorden?

Als bureauonderzoeker kun je met een snelle scan op de bovenstaande indicatoren het databestand substantieel verbeteren. Een aantal dataleveranciers beschikt over geautomatiseerde tools waarmee je controles kunt uitvoeren. Als opdrachtgever van een marktonderzoek krijg je niet altijd zelf toegang tot de ruwe data van een onderzoek en kun je dus weinig invloed uitoefenen op de controle van respondenten. Wel kun je actief vragen naar welke datacontroles zijn uitgevoerd door het bureau en eventueel de dataset opvragen om zelf te zoeken naar niet-serieuze respondenten.

Veel problemen achteraf kunnen vooraf worden voorkomen, namelijk door een goede vragenlijst op te stellen. Saaie, onduidelijke, irrelevante of lange vragenlijsten voorleggen is vragen om frauderende respondenten. Kun je echt verwachten dat een respondent zorgvuldig een statementbatterij van 20 of meer statements bij langs gaat?

Als onderzoeker moet je rekening houden met de beperkingen van het brein van je respondenten. Mensen kunnen en willen niet alles beantwoorden wat we ze voorleggen, maar tegelijkertijd dwingen we ze om iets in te vullen. We gaan in deze paper verder niet in op het maken van een goede vragenlijst, maar we constateren wel dat dat het tegenwoordig technisch gezien een stuk eenvoudiger is om een onderzoek op te zetten. Dit heeft er toe geleid dat de aandacht voor het opstellen van een goede vragenlijst sterk verminderd is, wat het probleem van frauduleuze respondenten verergert.

Steekproeftrekking in de praktijk van kwantitatief onderzoek

Een volledig a-selecte personensteekproef in onderzoek met ondervraging is onhaalbaar. Dit komt met name door de non-respons, het niet in staat of bereid zijn om mee te werken aan onderzoek. Merk op dat bij observatie en bij onderzoek onder andere elementen dan personen het wel goed mogelijk is een volledig a-selecte steekproef te trekken. Ook bij specifieke doelgroepen - bijvoorbeeld leden van een vereniging - kan met een goede manier van benaderen - en daarmee een hoge respons - de a-selecte steekproef goed benaderd worden.

De vaststelling dat een volledig a-selecte personensteekproef bij ondervraging praktisch niet of nauwelijks meer te realiseren is, betekent niet dat de gedachtegang van de a-selecte aanpak geheel los gelaten kan worden. De gehele statistiek rondom het berekenen van nauwkeurigheden en het toetsen van onderzoeksresultaten is gebaseerd op de basisaanname van een volledig a-selecte steekproef. Een onderzoeker doet er goed aan te streven naar een zo a-select mogelijke steekproef en/of naar representativiteit op basis van bekende kenmerken. Dit betekent aandacht voor het steekproefkader en voor het realiseren van een hoge respons. Het simpelweg verhogen van het aantal te benaderen personen (de bruto steekproef) om toch maar het gewenste aantal respondenten te krijgen, is geen goede oplossing. Een eventuele vertekening door de manier van steekproeftrekken of door het steekproefkader wordt hiermee niet opgelost. Daarnaast moet achteraf de gerealiseerde (netto) steekproef met behulp van de bekende kenmerken getoetst worden om te bekijken of de netto steekproef niet afwijkt van de populatie. In het geval dat er wel afwijkingen worden vastgesteld, zal moeten worden herwogen.

Maar er zijn uitzonderingen. Als je concludeert dat het niet strikt noodzakelijk is om de uitkomsten van je onderzoek te kunnen generaliseren over de hele doelgroep, dan kun je afwijken van de eisen die aan de representativiteit van de steekproef worden gesteld. Dit kan het geval zijn als je onderzoeksvragen exploratief en/of hypothesevormend van aard zijn. Denk aan:

- Het genereren van zoveel mogelijk ideeën voor nieuwe producten, nieuwe smaken of verbeteringen aan bestaande producten;
- Het opdoen van inspiratie voor het ontwikkelen van nieuwe reclamecampagnes;
- Het vinden van mogelijke verklaringen voor bepaalde ontwikkelingen in de markt, zoals een afnemend marktaandeel.

Voor deze vraagstellingen is helemaal geen representatieve steekproef nodig. Sterker nog, je wilt misschien juist een selecte groep van deelnemers aan je onderzoek. Denk bijvoorbeeld aan creatievere consumenten of uitgesproken fans van je merk.

Steekproeftrekking bij kwalitatief onderzoek

Bij kwalitatief onderzoek spreken we meestal over de 'selectie' van respondenten. Maar eigenlijk is dit niets anders dan een steekproef. We doen onderzoek onder een deel om op basis daarvan iets te kunnen zeggen over het geheel. Dat 'iets' wordt daarbij niet in kwantitatieve termen geformuleerd (*hoeveel*), maar is meer gericht op het *wat of hoe*. Omdat er in kwantitatieve zin geen projectie gemaakt wordt van steekproef naar populatie, hoeft niet te worden voldaan aan de eis van representativiteit. Wel is het zaak om ervoor te zorgen dat het steekproefkader zo goed mogelijk de doelgroep afdekt. Ook is het aan te raden ervoor te zorgen dat subgroepen waarvan je mag aannemen dat ze verschillend denken over het onderwerp in de steekproef vertegenwoordigd zijn.

Ook voor kwalitatief onderzoek geldt dat je niet altijd een goede afspiegeling van de doelgroep wilt hebben. Soms wil je juist de voorlopers, degenen die anders denken. In veel gevallen kan het ook verhelderend werken om consumenten te selecteren waarvan je op voorhand weet dat ze het niet eens zijn met elkaar op de issues die je wilt onderzoeken. Vooral in groepsdiscussies kan dit leiden tot geanimeerde discussies waarin een veelheid van standpunten de revue passeert.

De toekomst: Social Media Sampling en Sampling door Google

Het kunnen vinden van de juiste respondenten en ze bereid vinden om hun mening te geven, is altijd één van de grootste struikelblokken in marktonderzoek geweest. Online onderzoek met behulp van respondentenpanels heeft het ons makkelijker gemaakt, hoewel we nu steeds meer met de nadelen ervan worden geconfronteerd. Een ontwikkeling die in de toekomst een belangrijke rol gaat spelen, is het inzetten van sociale media als steekproefinstrument. Wereldwijd zijn er bijna twee miljard mensen lid van Facebook, in Nederland meer dan 8 miljoen (volgens gegevens van Facebook zelf). Deze leden zijn in potentie te benaderen voor onderzoek en vormen daarmee een steekproefkader. Het is een kwestie van tijd tot de grote sociale media hun leden aanbieden als steekproefkader. Onderzoeksbureaus kunnen hieruit putten, zowel voor kwantitatief als (online) kwalitatief onderzoek.

Omdat consumenten zichzelf steeds meer blootgeven op sociale netwerksites, wordt het ook steeds makkelijker om de juiste consumenten te vinden om te rekruteren voor je onderzoek. Eigenaren van sociale netwerken kennen de demografische kenmerken van hun leden, maar ook hun interesses en online gedrag kunnen worden meegenomen bij het trekken van een steekproef. Het gedrag en de interesses kunnen bovendien worden gebruikt in de analysefase, waardoor veel krachtiger analyses gemaakt kunnen worden, zonder dat het nodig is om meer vragen te stellen.

In dit kader is een initiatief van Google interessant. Onder de titel Consumer Surveys² biedt het bedrijf de mogelijkheid om vragen te stellen aan bezoekers van websites. Daarnaast zijn er vele analysemogelijkheden op basis van achtergrondkenmerken van deze bezoekers die al bekend zijn bij Google. Volgens Google is men in staat om een steekproef te trekken van internetgebruikers die zeer dicht in de buurt komt van een a-selecte steekproef van internetgebruikers. De verwachting is dat Google hiermee op termijn een belangrijke concurrent zal worden van de huidige consumentenpanels. Overigens is het al enige stil rondom het initiatief

² Zie: www.google.com/insights/consumersurveys/how

Inzicht & Impact



Marketing intelligence

www.inzichtimpact.nl