



## Steekproeftrekking

In dit artikel worden twee begrippen beschreven die van belang zijn voor het uitvoeren van een onderzoek. Het gaat om de *populatie* van het onderzoek en de *steekproef*. Voor wat betreft steekproeven lichten we 10 methoden van steekproeftrekking toe (vijf probabilistische en vijf niet-probabilistische methoden) en zetten we uiteen hoe je een steekproefomvang kunt berekenen.

### Onderzoekspopulatie

Bij het uitvoeren van een onderzoek wil je meestal uitspraken kunnen doen die gelden voor een grote groep mensen of organisaties. De groep waarover je uitspraken wilt doen, noemen we de *populatie*. Bijvoorbeeld alle Nederlanders van 18 jaar en ouder. Of alle Nederlanders die in het bezit zijn van merk X. Of alle bedrijven die een vertegenwoordiger in dienst hebben. Het is belangrijk dat je – voordat je je onderzoek uitvoert – de populatie helder beschrijft en afbakt.

Alle gegevens die uit een populatie te halen zijn (zoals leeftijd, inkomen, opleidingsniveau, koopgedrag, etc.) noemen we het *universum* aan gegevens. Het is belangrijk dat een onderzoeker zich beperkt tot het achterhalen van gegevens uit het universum die relevant zijn voor het onderzoek.

### Steekproef

Voor de meeste onderzoeken geldt dat niet de gehele populatie wordt onderzocht, omdat dit te duur is en te veel tijd kost. Er wordt dan een zogeheten *steekproef* getrokken. Je onderzoekt dan een deel van de populatie, waarbij je na het onderzoek uitspraken wilt doen die gelden voor de hele populatie. Het is daarom belangrijk dat een steekproef *representatief* is voor de populatie. De meest representatieve steek-

proef is die, die even groot is als de populatie. Maar dan is het eigenlijk geen steekproef meer. Stel dat je onderzoekt hoeveel mensen in Nederland van 12 jaar en ouder dagelijks rookt, en dat je dit doet door het aan iedereen in de populatie te vragen (wat natuurlijk ondoenlijk is). En stel dat je vindt dat 25% van alle Nederlanders van 12 jaar en ouder dagelijks een sigaret opsteekt. We noemen dit gevonden resultaat de 'true value'. Eigenlijk is de 'true value' in onderzoek altijd een onbekende waarde, omdat het ondoenlijk is om iedereen uit de populatie in je steekproef mee te nemen. Omdat een steekproef (veel) kleiner is dan de populatie, is de kans groot dat de in de steekproef gevonden waarde iets afwijkt van de 'true value' (in de steekproef vind je bijvoorbeeld dat 24% rookt). In het algemeen geldt: hoe kleiner de steekproef, hoe groter de kans dat de steekproefwaarde afwijkt van de 'true value'. Andersom geredeneerd: hoe groter de steekproef, des te groter de kans dat de steekproefwaarde de 'true value' benadert (dit noemen we in de statistiek de 'centrale limietstelling'). Als twee verschillende steekproeven naar dezelfde vraag en in dezelfde populatie geheel andere uitkomsten geven, ligt dat hoogstwaarschijnlijk aan te kleine steekproeven.

### Steekproefmethoden

Er zijn diverse manieren om een steekproef te trekken. Er zijn *probabilistische* (aselecte) en *niet-probabilistische* methoden. Bij de probabilistische methode heeft elk element een bekende kans (niet gelijk aan 0) om in de steekproef terecht te komen. Dit is niet het geval bij een niet-probabilistische methode. Bij probabilistische steekproeven is het daarom mogelijk om de betrouwbaarheid en nauwkeurigheid in te schatten. In onderstaand kader worden in totaal tien methoden van steekproeftrekking toegelicht.

#### *Probabilistische methoden*

- *(Enkelvoudig) aselect*: elk element in de populatie heeft bij deze methode een gelijke kans om in de steekproef terecht te komen (er zijn dus geen restricties). Zoals bij een lottotrekking.
- *Systematisch*: op vaste punten (intervallen) wordt telkens één element uit de populatie geselecteerd. Stel dat populatie bestaat uit 1.500 studenten en je wilt een steekproefgrootte van 300 studenten. De steekproeffractie is dan  $1.500/300 = 1/5^e$ . Als je vervolgens een getal tussen 1 en 5 kiest (bijvoorbeeld 4), dan worden de volgende elementen in de steekproef opgenomen: 4, 9, 14, 19, 24, ..., 1499.
- *Gestratificeerd enkelvoudig aselect*: de populatie wordt eerst verdeeld in een aantal groepen ('strata'). Bijvoorbeeld grote en kleine bedrijven of gezinsgrootte. Uit deze groepen worden vervolgens aselect elementen getrokken. Dit kan proportioneel of disproportioneel. Bij proportioneel wordt rekening gehouden met de verdeling van de strata in de populatie en bij disproportioneel wordt dit niet gedaan.

- *Cluster*: in de populatie worden een aantal groepen, en vervolgens aselekt een aantal subgroepen geselecteerd. Uit deze subgroepen worden alle elementen in de steekproef opgenomen. Bijvoorbeeld:
  - Groepen: selectie van universiteiten;
  - Elementen: alle eerstejaars die bij de geselecteerde universiteiten staan ingeschreven.
- *Getrapt*: hier geldt hetzelfde als bij de clustersteekproef. Echter, in plaats van dat alle elementen uit de subgroep in de steekproef worden opgenomen (alle eerstejaars studenten), wordt hieruit een aselekte steekproef getrokken.

#### *Niet-probabilistische methoden*

- *Convenience sample*: de steekproef wordt getrokken op basis van gemakscriteria. Bijvoorbeeld: je zet een vragenlijst uit in je LinkedIn-netwerk. De methode is een serieuze bedreiging voor de validiteit van je onderzoek.
- *Beoordelingssteekproef*: dit is een zogenaamde doelgerichte steekproef en hoort in principe thuis in de exploratieve fase van een onderzoek (wordt veel gebruikt voor diepte-interviews en groepsdiscussies). In principe wordt er gebruik gemaakt van gemaksoverwegingen om de elementen te selecteren, maar ze dienen wel aan bepaalde voorwaarden te voldoen. Bijvoorbeeld als je interviews met zware rokers wilt afnemen. Mensen die niet roken worden dan niet in de steekproef opgenomen.
- *Random walk*: deze methode wordt ingezet voor huis-aan-huis onderzoek. Eerst wordt willekeurig een startadres bepaald. Vanuit dit adres wordt een vast criterium gegeven om het volgende adres te selecteren (bijvoorbeeld: elk vijfde huis na het laatst bezochte adres).
- *Sneeuwbal*: deze methode wordt toegepast wanneer elementen in een populatie moeilijk te bereiken zijn. Zoals directeurs van grote bedrijven. Tijdens een interview met bekende leden van de groep kan gevraagd worden of zij de gegevens van andere leden uit de groep kunnen verstrekken, enzovoorts. Zo wordt de steekproef steeds groter.
- *Quota*: er wordt voor gezorgd dat de steekproef op van tevoren bepaalde kenmerken (denk hierbij aan leeftijd, inkomen, etc.) dezelfde verdeling heeft als de populatie. Deze methode is niet aselekt.

### Steekproefomvang

Een steekproef dient groot genoeg te zijn om voor de populatie representatieve uitspraken te kunnen doen. Om de omvang van de steekproef ( $n$ ) te kunnen bepalen moeten een aantal stappen worden ondernomen:

1. De omvang van de populatie ( $N$ ) moet worden bepaald. De omvang van de populatie is een extern gegeven en is niet beïnvloedbaar door de onderzoeker. Zoals het aantal inwoners van Nederland van 18 jaar en ouder.
2. Het bepalen van de mate waarin het te onderzoeken gegeven in de populatie voorkomt (uitgedrukt in een percentage). Deze mate is meestal niet bekend en dient geschat te worden. Dit kan gebeuren op basis van eerder onderzoek of intuïtie. Als het percentage niet is in te schatten, dient voor de zekerheid 50% aangehouden te worden. Dit levert de grootst mogelijke steekproefomvang op en sluit zo zoveel mogelijk risico's uit.
3. Bepaal de gewenste (on)nauwkeurigheid. Dit percentage geeft de mate aan waarbinnen je uitspraken wilt doen. Meestal gaan we uit van een nauwkeurigheid die tussen de 1 – 5% ligt (hoe kleiner het percentage hoe groter de nauwkeurigheid). Na punt (4) komen we hier op terug.
4. Bepaal de gewenste betrouwbaarheid waarover je uitspraken wilt doen. 95% betrouwbaarheid wordt het meest gebruikt, maar ook 90% en zelfs 99% komen voor. Bij deze percentages hoort een bepaalde Z-score:
  - 90% betrouwbaarheid → Z-score = 1,65
  - 95% betrouwbaarheid → Z-score = 1,96
  - 99% betrouwbaarheid → Z-score = 2,58

Hoe moet je de gewenste (on)nauwkeurigheid en de gewenste betrouwbaarheid nu interpreteren? Stel dat we in een steekproef vinden dat 24% van alle Nederlanders van 12 jaar en ouder dagelijks rookt. En dat we 90% zekerheid daar uitspraken over willen doen met een (on)nauwkeurigheid van 2%. Formeel kunnen we dan concluderen dat:

*We met 90% zekerheid kunnen zeggen dat het percentage rokers in de populatie tussen de 22 en 26% ligt (i.e. 24 min 2% en plus 2%).*

Anders geformuleerd:

*Als we 100 steekproeven trekken in dezelfde populatie, zal de waarde van de uitkomst bij 90% van deze steekproeven liggen tussen de 22 en 26%.*

Willen we met een hogere zekerheid (betrouwbaarheid) uitspraken doen, dan betekent dit dat – bij dezelfde steekproefgrootte – de onnauwkeurigheidsmarge toeneemt. Als we de onnauwkeurigheidsmarge hetzelfde willen houden, betekent dat dat we de steekproefgrootte moeten verhogen. Hieronder leggen we uit hoe je de steekproefgrootte kunt berekenen.

Als je een steekproefgrootte wilt berekenen, moet je allereerst de grootte van de te onderzoeken populatie inschatten. Als deze groter is dan 20.000, voer dan onderstaande berekening uit. Is de populatie kleiner dan 20.000, bereken dan  $n'$ , de *gecorrigeerde steekproefomvang* (dit wordt verderop in dit document toegelicht).

### Berekenen van een steekproefgrootte

Bij het berekenen van een steekproefgrootte, zijn twee formules relevant:

$$\text{onn.} = Z \cdot S_x \quad (a)$$

$$S_x = \sqrt{\frac{P \cdot Q}{n}} \quad (b)$$

Waarbij:

- onn. : gekozen onnauwkeurigheid (zie punt 3 hierboven);
- Z : gewenste betrouwbaarheid (zie punt 4 hierboven);
- $S_x$  : standaardmeetfout;
- P : mate waarin het te onderzoeken gegeven in de populatie voorkomt (zie punt 2 hierboven);
- Q :  $Q = (100 - P)$ ; ofwel: Q is de inverse van P.

Stel, dat je geen idee hebt van de mate waarin het te onderzoeken gegeven in de populatie zich manifesteert (P). Hierboven (onder punt 2) gaven we aan dat je dan P op 50% moet stellen en Q is dan ook 50%. En stel dat we een onnauwkeurigheidsmarge kiezen van 5% (onn. = 5) en dat we de gewenste betrouwbaarheid op 95% stellen ( $Z = 1,96$ ). We kunnen dan met formule (a) de standaardmeetfout ( $S_x$ ) berekenen:

$$5 = 1,96 \cdot S_x$$

ofwel:

$$S_x = 5 / 1,96 = 2,55$$

Omdat we de waarde van de standaardmeetfout ( $S_x$ ) nu weten en ook die van P en Q, kunnen we formule (b) invullen:

$$2,55 = \sqrt{\frac{50 \cdot 50}{n}}$$

$$2,55 = \sqrt{\frac{2.500}{n}}$$

$$2,55^2 = \frac{2.500}{n}$$

$$6,51 \cdot n = 2.500$$

$$n = 384$$

De benodigde steekproef bij 95% betrouwbaarheid en een onnauwkeurigheid van 5% is 384, als het element waar je uitspraken over wilt doen in 50% van de gevallen voorkomt. Als deze laatste verwachting kleiner of groter is, dan wordt de benodigde steekproef kleiner. Let er op dat het hier om de te *behalen* steekproefgrootte gaat. Als de respons op een enquête 10% is, betekent dit dus dat je  $10 \times 384 = 3.840$  enquêtes moet uitzetten.

Het is tevens mogelijk om de betrouwbaarheid te berekenen op basis van de behaalde respons in een onderzoek. Stel dat je met de hierboven genoemde uitgangspunten een respons op je enquête hebt behaald van 200. Welk betrouwbaarheidsniveau hoort hier dan bij? We vullen eerst formule (b) in:

$$S_x = \sqrt{\frac{50 \cdot 50}{200}}$$

$$S_x = \sqrt{12,5} = 3,54$$

Vervolgens formule (a):

$$\text{onn.} = Z \cdot S_x$$

$$5 = Z \cdot 3,54$$

$$Z = 1,41$$

De Z-score (1,41) is kleiner dan 1,65. Dat wil zeggen dat op basis van de bovenstaande gegevens de betrouwbaarheid kleiner is dan 90%.

**Gecorrigeerde steekproefomvang**

Als een populatie kleiner is dan 20.000 elementen, verdient het aanbeveling om een gecorrigeerde steekproefomvang te berekenen. Dit doe je door eerst een 'gewone' steekproefomvang uit te rekenen (zoals hierboven waar we een gewenste steekproefomvang van 384 vaststelden). Als de populatie minder dan 20.000 elementen telt, kun je aan de hand van formule (c) de gecorrigeerde steekproefomvang ( $n'$ ) berekenen:

$$n' = \frac{n}{1 + (n/N)} \quad (c)$$

Waarbij:  $n'$  : gecorrigeerde steekproefomvang;  
 $n$  : berekende steekproefomvang conform formules (a) en (b)  
 $N$  : grootte van de populatie.

Stel dat de populatie 1.000 elementen telt. De gecorrigeerde steekproefomvang bedraagt dan:

$$n' = \frac{384}{1 + (384/1.000)}$$

$$n' = 277$$

Bij een populatie van 1.000 elementen bedraagt de gecorrigeerde steekproefomvang dus 277. Als je de gecorrigeerde steekproefomvang berekent voor een nog kleinere populatie ( $N = 500$ ), dan bedraagt deze 217. Bij 1.000 elementen moet je dus 27,7% van die elementen opnemen in je steekproef en bij 500 is dit gestegen tot 43,4%. Met andere woorden: hoe kleiner de populatie, des te groter de steekproef in verhouding tot die populatie moet zijn. Bij een populatie van 100 elementen moet je zelfs 79,3% van het aantal elementen van de populatie in je steekproef meenemen (i.e. 79 elementen). Nogmaals: je berekent de gecorrigeerde steekproefomvang alleen als je populatie (naar schatting) kleiner is dan 20.000 elementen.

*Referentie(s)*

Slotboom, A. (1987), Statistiek in woorden: De meest voorkomende termen en technieken. Wolters-Noordhoff, Groningen.

Baarda, D.B., Goede, M.P.M. de (2006), Basisboek methoden en technieken. Wolters-Noordhoff, Groningen/Houten.