Inhoud

Inleiding tot de statistiek

1

In dit hoofdstuk leer je de betekenis van de volgende veelvoorkomende statistische termen:

- **beschrijvende statistiek**
- **verklarende statistiek**
- **populatie**
- **steekproef**
- **representatieve steekproef**
- **aselect**
- **variabele**
- **parameter**
- **schalen (nominale, ordinale, interval- en ratioschaal)**

1.1 Beschrijvende en verklarende statistiek

Bijna 6000 jaar geleden waren de Babyloniërs al met statistiek bezig. Iedere zes à zeven jaar werden alle Babyloniërs geteld, samen met hun veestapel en hun agrarische productie. Waarom deden zij dit?

Er zijn verschillende redenen waarom de Babyloniërs deze volkstellingen hielden. De meest aannemelijke verklaring is dat ze er zeker van wilden zijn dat er altijd genoeg voedsel was voor iedereen. Daarom was het van belang om gebieden met tekorten of overschotten snel te kunnen opsporen. Dit konden ze pas goed doen als ze niet alleen telden, maar ook hun gegevens goed ordenden en noteerden. Die gegevens (mensen, dieren en productie) werden keurig per gebied gerangschikt en in tabellen gezet.

De Babyloniërs waren bezig met wat we nu beschrijvende statistiek noemen. De drie belangrijkste acties binnen de beschrijvende statistiek zijn: het verzamelen, het organiseren en het presenteren van gegevens.

Tegenwoordig worden in veel landen nog steeds volkstellingen gehouden. Deze volkstellingen gaan verder dan alleen het tellen van mensen en dieren; er worden onder andere vragen gesteld over leeftijd, godsdienst, samenlevingsvormen en onderwijs. Al deze informatie wordt verzameld, geordend en gepubliceerd door statistische bureaus van deze landen. De resultaten hebben onder andere invloed op de toewijzing van subsidies en op het openbare beleid.

Het verwerken van volkstellingen wordt nog altijd voornamelijk verricht met behulp van de beschrijvende statistiek. Maar tegenwoordig doen we veel meer met statistiek dan alleen beschrijven. Het analyseren en interpreteren van gegevens is steeds belangrijker geworden. En dat is het terrein van de verklarende statistiek.

Veel moderne volkstellingen gaan vergezeld van een gedetailleerde vragenlijst, die slechts door een klein gedeelte van de bevolking hoeft te worden ingevuld. Deze geselecteerde groep mensen noemen we een steekproef. De vragenlijst kan gaan over zaken als gezondheid of politieke voorkeur. Op basis van de resultaten van de steekproef worden uitspraken gedaan over de gehele populatie. Met andere woorden: de verzamelde gegevens worden geanalyseerd en geïnterpreteerd.

In Nederland worden helemaal geen volkstellingen meer gehouden. Populatiegegevens worden verzameld via het bevolkingsregister. Verder worden veel steekproeven genomen door instanties als het Centraal Bureau voor de Statistiek (CBS). De resultaten van deze steekproeven zijn onder andere te vinden in een statistisch jaarboek, dat ieder jaar door het CBS wordt uitgegeven en gedeeltelijk ook op het internet verschijnt (www.cbs.nl).

Het verschil tussen beschrijvende en verklarende statistiek wordt voornamelijk bepaald door het gebruik van steekproeven.

Zoals gezegd bestaat de beschrijvende statistiek uit het verzamelen, ordenen en presenteren van gegevens. Dit gebeurt vaak op basis van populatiegegevens, maar dat hoeft natuurlijk niet. Ook gegevens die je hebt verzameld door middel van een steekproef, kun je ordenen en presenteren.

De verklarende statistiek werkt altijd met steekproeven. Deze steekproeven worden gebruikt om uitspraken over populaties te doen. Dit zijn nooit exacte uitspraken, zoals: 'de Nederlandse vrouw is gemiddeld 166,7896364763498 cm lang'. Dat is niet mogelijk, want je hebt nu eenmaal niet alle gegevens van een hele populatie. Toch kun je de werkelijkheid vaak heel goed benaderen met uitspraken als: 'de Nederlandse vrouw is gemiddeld ongeveer 167 cm lang' of 'met 95% betrouwbaarheid bevat het interval van 165 cm tot 169 cm de werkelijke gemiddelde lengte van Nederlandse vrouwen'.

Er zullen altijd fluctuaties optreden tussen de resultaten van verschillende steekproeven uit één populatie. Soms zijn deze fluctuaties toevallig, soms zijn ze niet toevallig maar is de oorzaak (nog) onbekend. Bij nadere beschouwing kunnen factoren worden gevonden die een deel van deze fluctuaties verklaren. Met andere woorden: de gegevens, verkregen door steekproeven, moeten worden geanalyseerd en geïnterpreteerd vóór je een betrouwbare uitspraak kunt doen over de populatie. Bijna alle moderne statistiek is verklarende statistiek.

1.2 Populatie en steekproef

Iedereen die onderzoek doet, krijgt te maken met statistiek. Dit kan diepgaand wetenschappelijk onderzoek zijn, maar ook een klein onderzoekje waarbij je bijvoorbeeld op straat wordt gevraagd wat je de lekkerste koekjes vindt.

Vaak is het bedrijven van statistiek, hoe belangrijk ook, niet het hoofddoel van het onderzoek. Meestal staat het ten dienste van een ander vakgebied. Neem nou de koekjesfabriek Hapslik. De directie wil weten wat consumenten van hun koekjes vinden. Ze hebben statistiek nodig om hierover valide uitspraken te kunnen doen. Daarom is statistiek een hulpwetenschap. Slechts een klein onderdeel van de statistiek betreft puur statistisch onderzoek.

Bij het meeste onderzoek gaat het erom uitspraken te doen over een bepaalde populatie. Het begrip populatie betekent in de statistiek dan ook meer dan 'alle inwoners van een land'. Het is 'een verzameling elementen'. Deze elementen kunnen mensen zijn, maar ook andere dingen die je wilt onderzoeken, bijvoorbeeld dieren of gebeurtenissen. Dus ook 'alle mussen in een park' vormen een populatie; of 'alle koekjes gemaakt in één fabriek'; of zelfs 'alle operaties in een ziekenhuis in één jaar'. Essentieel is dat het om een groep elementen gaat waarvan je één of enkele kenmerken wilt onderzoeken. De directie van de koekjesfabriek Hapslik wil van de populatie potentiële Hapslik-koekjeskopers weten wat zij van de smaak van de Hapslik-koekjes vindt.

Een steekproef is een deelverzameling van de populatie. Een steekproef wordt genomen om iets te kunnen zeggen over een populatie, zonder daarvoor de hele populatie te hoeven onderzoeken. Als jou op straat wordt gevraagd wat jij van koekjes van het merk Hapslik vindt, dan behoor je tot de steekproef uit de populatie potentiële Hapslik-koekjeskopers. De directie van de fabriek zal op basis van deze steekproef conclusies trekken over de hele populatie potentiële Hapslik-koekjeskopers.

De bedoeling van een steekproef is om zo veel mogelijk informatie over een populatie te krijgen met zo min mogelijk metingen. Meestal worden steekproeven dan ook genomen om *tijd en geld te besparen*.

Als men vlak voor de verkiezingen een idee wil krijgen over het stemgedrag van de Nederlanders, is het veel te omslachtig en te duur om iedere stemgerechtigde te vragen wat hij of zij denkt te gaan stemmen. Daarom wordt slechts een gedeelte van de stemgerechtigden gepeild. Dat is goedkoper én sneller.

Soms worden steekproeven uitgevoerd omdat het *fysiek onmogelijk* is een hele populatie te meten. Probeer maar eens alle vissen in de zee te tellen!

Het kan ook voorkomen dat een *populatie oneindig* is. Het gooien van een dobbelsteen kan oneindig vaak gebeuren. Om er een idee te van krijgen of alle zijden even vaak boven komen – en zo niet, welke zijden dan vaker boven komen – kun je een steekproef nemen door bijvoorbeeld honderd keer met de dobbelsteen te gooien.

Een steekproef is ook nuttig als je iets wilt meten dat *kapot kan gaan* door deze meting. Als je in een touwfabriek wilt testen bij welke kracht een bepaald type touw knapt, is het niet slim om net zo lang aan alle touwen te trekken totdat ze kapotgaan. Je weet dan bij welke kracht het touw gemiddeld knapt, maar je hebt niets meer over. Voer je de krachttest uit op een steekproef uit al het geproduceerde touw, dan kun je de kracht schatten waarbij dit type touw gemiddeld knapt.

Naast dit alles vermindert het gebruik van steekproeven het aantal meetfouten. In een steekproef worden minder metingen gedaan dan in een hele populatie. Zo kunnen er ook minder fouten worden gemaakt.

Een steekproef kan alleen goede en voldoende informatie over een populatie geven als de steekproef representatief is voor deze populatie. Wat is volgens jou een representatieve steekproef?

Als we het kiesgedrag van de Nederlanders willen peilen, vorm jij dan een representatieve steekproef?

Nee, ik vrees van niet. Als jij samen met nog 999 anderen was uitgekozen, dan was het een ander verhaal, maar één persoon is niet genoeg. Als jij PvdA stemt, zegt dat nog niets over de rest van de bevolking. Hoe groter de steekproef, des te beter beeld je van de populatie krijgt. Aan de andere kant is een kleinere steekproef sneller en goedkoper. Het is altijd de kunst om een goed evenwicht te vinden tussen deze twee uitersten.

Zou een schoolklas een representatieve steekproef kunnen vormen voor het kiesgedrag van de Nederlanders?

Een schoolklas is natuurlijk te klein voor deze representatieve steekproef, maar dat is niet het belangrijkste. De meeste schoolkinderen zijn jonger dan 18 en hebben niet de stemgerechtigde leeftijd. Ze maken dus geen deel uit van de doelpopulatie van deze steekproef. De doelpopulatie betreft hier namelijk niet 'de Nederlanders', maar 'de stemgerechtigde Nederlanders'. De meeste schoolkinderen mogen nog niet stemmen, dus kan een schoolklas nooit een representatieve steekproef vormen.

Wat dan te denken van alle volwassen bezoekers aan een golfbaan op één dag?

In Nederland is golf een redelijk elitaire sport. Waarschijnlijk geeft een peiling van het stemgedrag van golfers daarom een scheef beeld. Het is dus ook belangrijk dat je voor je steekproef niet een specifieke groep uitkiest, maar juist zo willekeurig mogelijk te werk gaat.

Het toeval speelt hierbij een belangrijke rol. Je kunt bijvoorbeeld honderd willekeurige stemgerechtigden kiezen uit het bevolkingsregister van tien willekeurig gekozen gemeenten. In statistische termen is zo'n steekproef aselect. Dat wil zeggen louter door het toeval bepaald en dus geheel willekeurig.

1.3 Random number generator

De meest pure manier om een willekeurige, oftewel aselecte, steekproef te nemen is met behulp van tabellen met aselecte getallen. Meestal gebruiken we hiervoor de Engelse term random number tables.

De random number tables zijn overzichten van volledig willekeurige getallen die bijvoorbeeld in kolommen van vijf cijfers worden weergegeven, zoals in de volgende tabel:

64894	74296	24805	24037	20636
19645	93030	23209	25600	90376
70715	12987	53985	11298	76585
80157	36147	64032	36653	98951
34072	76850	35697	31760	65813
45571	82406	39533	30335	70225
66347	95327	56298	40041	40690
15325	47048	90553	57348	28463
30529	64778	35808	34282	60935
60098	37588	80456	90756	25678
41941	50949	19435	48581	88695

Voor je deze tabel kunt gebruiken, moet je eerst de hele populatie waaruit je de steekproef wilt nemen, nummeren. Dus in het geval van de Nederlandse stemgerechtigden krijgt iedereen een nummer van 1 tot 12 miljoen, of wat dan ook het exacte aantal stemgerechtigde Nederlanders is.

Om een niet al te langdradig voorbeeld te geven gaan we ervan uit dat de populatie alleen de 800 stemgerechtigde inwoners van een dorp betreft. Stel, je wilt een idee van het stemgedrag van het dorp krijgen. Met behulp van een random number table neem je een aselecte steekproef van 20 dorpelingen. Je nummert alle 800 inwoners. Dan kies je op een willekeurige plaats in de tabel het eerste aselecte getal. Je kunt bijvoorbeeld je ogen dichtdoen en gewoon prikken. Of je kiest – als je op 2 mei jarig bent – het getal in de tweede kolom en de vijfde rij.

Vóór je verder gaat, moet je een aantal besluiten nemen. Allereerst is het nummer van elk van de 800 inwoners maximaal drie cijfers groot. Je kolommen bestaan uit vijf cijfers. Het heeft dus geen zin om alle cijfers uit de kolom te gebruiken. Je moet van tevoren beslissen of je steeds de eerste drie, de laatste drie of een andere combinatie van drie uit vijf gebruikt. Daarnaast moet je beslissen hoe je verder gaat nadat je het eerste aselecte getal hebt getrokken. Ga je in de tabel naar beneden, naar boven, naar links of naar rechts? Neem je steeds het volgende getal of sla je één of twee getallen over?

Stel, je kiest voor de eerste drie van de vijf cijfers. Voor ieder volgend getal ga je naar beneden en sla je steeds één rij over. Als je aan het eind van de kolom komt, ga je verder met de kolom die er rechts naast ligt. Een aselect getal dat groter is dan de populatie, in dit geval groter dan 800, moet worden overgeslagen. Niemand heeft dit nummer.

In dit voorbeeld zijn de eerste vijf getallen: 768, 647, 509, 232 en 640.

Dezelfde procedure kun je uitvoeren met een steekproef van 1000 uit de 12 miljoen stemgerechtigde Nederlanders. We raden je dan wel aan om gebruik te maken van een zogenaamde random number generator. Dit is een computerprogramma dat aselecte getallen uitspuwt op basis van een populatie- en steekproef-

grootte die je opgeeft. Het is te vinden op internet. Op de internetsite van deze cursus hebben we er ook een toegevoegd.

1.4 Soorten steekproeven

In plaats van een gewone aselecte steekproef kun je ook een systematische aselecte steekproef nemen. In dat geval hoef je maar één aselect cijfer te trekken. Dit kun je doen door je populatie op te delen in net zoveel groepen als de steekproef groot is. Elke groep bevat zo evenveel nummers, namelijk de populatiegrootte (N, bijv. 800) gedeeld door de steekproefgrootte (n, bijv. 20). Dus in het dorp uit de vorige paragraaf bestaat de eerste groep uit de inwoners met de nummers 1 tot 800/20 = 40 (k). De tweede groep bestaat uit de volgende veertig inwoners, namelijk de nummers 41 tot en met 80, enzovoort. Met behulp van een aselecte cijfertabel (of eventueel een generator) trek je een aselect getal tussen 1 en 40, bijvoorbeeld 24. Je steekproef van 20 bestaat dan uit de inwoners met de nummers: 24, 24 + 40 = 64, 24 + 40 + 40 = 104, enzovoort. Zo ga je door tot en met nummer 784.

Het trekken van een systematisch aselecte steekproef kan alleen als de nummering van de elementen in de populatie volstrekt willekeurig is (dus niet op basis van bijvoorbeeld leeftijd of woonwijk).

Naast de normale aselecte steekproef en de systematische aselecte steekproef bestaan er nog vele andere vormen van min of meer aselecte steekproeven. We zullen er hier een paar bespreken.

Allereerst is er de gelede (of gestratificeerde) steekproef. Deze wordt gebruikt als er duidelijke verschillen zijn tussen bepaalde groepen in de populatie. Niet alle bevolkingsgroepen vertonen hetzelfde stemgedrag. Jongeren stemmen anders dan ouderen, en die stemmen weer anders dan mensen met een leeftijd ertussenin. Ook gelovigen stemmen vaak anders dan mensen die niet geloven. Om een goed beeld van de hele populatie te krijgen moet je zorgen dat de samenstelling van je steekproef min of meer overeenkomt met de werkelijke samenstelling van de bevolking. Met andere woorden: als de ene groep 30% van de bevolking betreft en de andere groep 70%, moet je ervoor zorgen dat je steekproef ook ongeveer deze verhouding heeft. Dus selecteer je 30% van je steekproef uit de eerste groep en 70% uit de tweede groep. In een gelede steekproef bepaal je per groep het stemgedrag.

Naast de gelede steekproef kennen we ook de clustersteekproef. Ging de gelede steekproef uit van duidelijk aanwezige groepen, bij de clustersteekproef splitst een onderzoeker de populatie op in groepen. Een bekende methode is het opdelen van het land in vele kleine geografische gebieden. Je selecteert uit deze gebieden een aantal representatieve gebieden en ondervraagt alle inwoners in die uitgekozen gebieden.

Een variant op de clustersteekproef is de meerstadiasteekproef. Hierbij wordt de populatie niet alleen in groepen opgedeeld, maar worden er ook aselecte steekproeven genomen in een paar van de aselect gekozen groepen. Dus slechts een klein groepje mensen uit de gekozen gebieden wordt ondervraagd.

OPGAVE

VERDIEPING

1.5 Variabelen en parameters

Je weet inmiddels dat het lang niet altijd nodig is alle elementen van een populatie te meten om nauwkeurige informatie over deze populatie te verkrijgen. Een steekproef is vaak voldoende. Ook hoef je niet alle kenmerken van de elementen te meten. Je kunt je beperken tot de kenmerken waarin je geïnteresseerd bent.

Inwoners van Nederland hebben vele kenmerken. Een paar daarvan zijn: inkomen, lengte, gewicht, kleur ogen en kiesgedrag. In het ene onderzoek ligt de nadruk op kiesgedrag en voor een ander onderzoek ben je misschien meer geïnteresseerd in lengte en oogkleur.

In de statistiek noem je het kenmerk (of de kenmerken) waarnaar je onderzoek doet de variabele(n). Zo gaat je interesse in het ene onderzoek uit naar de variabele 'kiesgedrag', in een ander onderzoek naar de variabelen 'lengte' en 'oogkleur'.

Een variabele, het woord zegt het al, varieert. Zij kan allerlei verschillende waarden aannemen. Zo kunnen de lengtes van volwassen Nederlanders variëren tussen, zeg, 100 en 220 cm.

Wij zullen in alle hoofdstukken de naam van een variabele in cursieve hoofdletters weergeven. In het algemeen zullen we daarvoor de letter X gebruiken, en als er meer variabelen zijn de letters Y en Z. Metingen aan X, oftewel waarden van X, zullen we met kleine cursieve letters als x_i weergeven. De index 'i' geeft hierbij aan om welke meting het gaat. Vijf aselect gekozen Nederlanders hebben bijvoorbeeld de volgende lengtes (in cm):

$x_1 = 168$
$x_2 = 179$
$x_3 = 191$
$x_4 = 174$
$x_5 = 180$

Verwar variabelen, en zeker de metingen daaraan, niet met parameters. Parameters zijn vaste, vaak onbekende, populatiegrootheden. Zo is de gemiddelde lengte van alle Nederlanders een parameter. Deze gemiddelde lengte is (op één moment) een vast getal, al ken je de exacte waarde niet. We zullen parameters weergeven met Griekse letters. Zo zul je in hoofdstuk 3 zien dat we voor het populatiegemiddelde de Griekse mu (μ) gebruiken.

Zoals al eerder opgemerkt, wil je in de meeste onderzoeken uitspraken doen over populaties. Maar nu we hebben uitgelegd wat parameters zijn, kunnen we daarin nog exacter zijn. Eigenlijk doe je in de meeste onderzoeken uitspraken over populatieparameters. Bijvoorbeeld: in een onderzoek naar de lengte van Nederlanders ben je voornamelijk geïnteresseerd in de gemiddelde lengte van Nederlanders (eventueel per type bevolkingsgroep, geslacht of leeftijd). Om zonder de lengte van alle Nederlanders te hoeven meten tot een uitspraak over dit populatiegemiddelde te komen, kun je een aselecte steekproef van bijvoorbeeld 100 Nederlanders nemen. Dit zijn 100 waarden van de variabele lengte. Het steekproefgemiddelde, oftewel het gemiddelde van de 100 waarden, kun je gebruiken als schatter voor het populatiegemiddelde.

We willen nog een stap verder gaan in de wereld van variabelen en parameters. Om een aantal hoofdstukken hierna te kunnen begrijpen moet je weten dat er eigenlijk twee soorten variabelen zijn: de stochastische variabelen en de vaste variabelen. Beide soorten kunnen verschillende waarden aannemen, maar de waarden van een stochastische variabele zijn afhankelijk van het toeval, terwijl de waarden van een vaste variabele min of meer van tevoren zijn vastgesteld.

Stel dat je de relatie tussen de lengte en het gewicht van de Nederlanders wilt bepalen. Dit kan met behulp van een statistische techniek die regressieanalyse wordt genoemd (zie hoofdstuk 11). Hiervoor heb je een aselecte steekproef van je doelgroep nodig.

Deze steekproef kun je op verschillende manieren nemen.

Je kunt bijvoorbeeld 100 mensen aselect uit de bevolking trekken en hun lengte en gewicht meten. In dit geval zijn lengte en gewicht stochastische variabelen. De waarden van de lengtes en gewichten hangen helemaal af van welke mensen toevallig in je steekproef zitten. Je kunt ook een steekproef van 100 willekeurige Nederlanders nemen van wie 20 een lengte tussen 150 en 160 cm hebben, 20 een lengte tussen 160 en 170 cm, enzovoort. In dit geval is de lengte een vaste variabele. De lengte varieert nog steeds, maar de frequentie van bepaalde lengtes wordt niet bepaald door het toeval, maar door jou. Het gewicht van al deze mensen is nog steeds een stochastische variabele. De methode uit het laatste voorbeeld komt veel voor in experimenteel onderzoek; de keuze van de waarden van de vaste variabele(n) wordt bepaald met behulp van een zogenaamde statistische proefopzet.

1.6 Schalen

De waarden van variabelen zijn vaak een getal: Jan heeft een lengte van 185 cm en hij is 37 jaar. Soms is zo'n kwantificering niet mogelijk. Bijvoorbeeld als het om kwalitatieve variabelen gaat: Jan heeft groene ogen en blond haar. Je kunt deze indeling in kwantitatief en kwalitatief nog iets verfijnen met de volgende vier veelgebruikte schalen.

Nominale schaal

De waarden van nominale variabelen zijn puur kwalitatieve gegevens. Met deze gegevens kun je geen wiskundige berekeningen doen. Groene ogen zijn niet blauwe ogen + 2 of zo iets absurds. Groen is ook niet beter of slechter dan blauw.

Ordinale schaal

De waarden van ordinale variabelen zijn kwalitatieve gegevens die te ordenen zijn. Echte berekeningen kun je hier niet mee doen, maar je kunt wel een zekere ordening aangeven, bijvoorbeeld: wat is beter of wat komt eerder. De beste voorbeelden zijn de antwoordcategorieën in veel vragenlijsten. Hier kun je antwoorden geven als: volledig mee oneens, mee oneens, neutraal, mee eens, volledig mee eens.

Intervalschaal

De waarden van deze schaal zijn kwantitatieve gegevens. Verschillen tussen waarden van deze schaal hebben een betekenis. Je kunt optellen en aftrekken. Verhoudingen hebben geen betekenis. Je kunt niet vermenigvuldigen of delen. Dit heeft te maken met het feit dat variabelen op een intervalschaal geen vast nulpunt hebben. Het meest beroemde voorbeeld is de temperatuur. De uitspraak '30 graden minus 10 graden is 20 graden' is juist en betekenisvol, maar '40 graden is tweemaal zo warm als 20 graden' klopt niet. Dit kun je gemakkelijk zien als je de temperatuur in Fahrenheit uitdrukt in plaats van in Celcius: 40 °C = 104 F, en 20 °C = 68 F. Het moge duidelijk zijn dat 104 niet tweemaal 68 is. Dit komt doordat Fahrenheit een ander nulpunt heeft dan Celcius.

Ratioschaal

Waarden op deze schaal zijn puur kwantitatieve gegevens. Verschillen en verhoudingen hebben een betekenis. De meeste gegevens waarmee we in deze cursus werken, zullen van dit type zijn. Voorbeelden hiervan zijn de lengte en het gewicht van de Nederlanders.

Andere indelingen

Naast een indeling in kwantitatief en kwalitatief of in vier schalen kunnen we variabelen ook indelen in discrete en continue varia-

belen. Discrete variabelen zijn variabelen die alleen bepaalde gehele waarden kunnen aannemen en tussenliggende waarden overslaan. Zo kun je met een dobbelsteen alleen 1, 2, 3, 4, 5 of 6 gooien, maar niets ertussenin. Je kunt niet 4,5 of 4,78786 gooien. Continue variabelen kunnen alle waarden aannemen binnen een bepaald interval. Lengte is hiervan een voorbeeld. Mensen kunnen iedere lengte aannemen tussen 0 en – pak 'm beet – 250 cm.

Tot zover deze inleiding in de statistiek. Met de opgaven die bij dit hoofdstuk horen kun je op internet voor jezelf controleren of alles wat we tot nu toe hebben verteld duidelijk is, en of je klaar bent voor de wondere wereld van tabellen en figuren in hoofdstuk 2.

1.7 Samenvatting

In dit hoofdstuk heb je kennisgemaakt met een aantal basisbegrippen uit de statistiek. We hebben proberen duidelijk te maken dat veel statistische analyses zijn gebaseerd op steekproeven, waarbij een steekproef een deelverzameling is uit een populatie. Een populatie kan een groep mensen zijn (bijvoorbeeld alle inwoners van een dorp of alle Nederlanders), maar ook alle vissen in de oceaan of alle producten verkocht in 2008 door bedrijf X.

We hebben laten zien dat je steekproeven in allerlei soorten en maten hebt, en hebben ook omschreven waaraan een steekproef moet voldoen wil deze representatief zijn voor de populatie waaruit hij wordt getrokken.

Vervolgens hebben we uitgelegd wat het verschil is tussen variabelen en parameters en waarom deze begrippen zo belangrijk zijn in de statistiek.

Ten slotte hebben we laten zien dat je variabelen op grofweg vier niveaus kunt meten. Dan hebben we het over de nominale, ordinale, ratio- en intervalschaal.