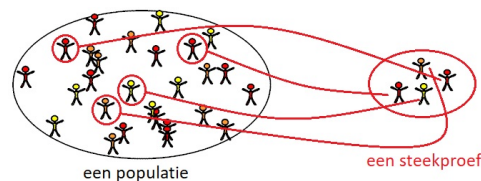


## 3.1 Steekproeven

### Inleiding

In het hoofdstuk 'Data verwerken' heb je geleerd hoe je gegeven data(sets) kunt presenteren, kunt samenvatten en kunt typeren.

Het verwerken van data is een belangrijk onderdeel van statistisch onderzoek. Maar er gaat iets aan vooraf. Het statistisch onderzoek moet wel een goede opzet hebben want anders zijn je antwoorden, je conclusies, niets waard. In dit onderdeel ga je vooral bekijken hoe je een goede steekproef moet opzetten.



Figuur 1

### Je leert in dit onderwerp

- de begrippen statistisch onderzoek, steekproef en populatie;
- wat een aselekt genomen en representatieve steekproef is;
- geschikte enquêtevragen opstellen.

### Voorkennis

- soorten statistische variabelen herkennen;
- werken met de normale verdeling.

### Verkennen

#### Opgave V1

In uitspraken in kranten, boeken en op internet kom je vaak resultaten van statistisch onderzoek tegen. Hier zie je daar een voorbeeld van.

Uit onderzoek van het Centraal Bureau voor de Statistiek (CBS) blijkt dat bijna de helft van de jongeren tussen de 15 en 25 jaar gebruik maakt van internet op de telefoon. Dat is veel meer dan vorig jaar, toen nog maar 20 procent van de jongeren internetten op hun mobiel.

(Bron: jongeren.blog.nl maart 2010)

Het Centraal Bureau voor de Statistiek (CBS) heeft zich kennelijk afgevraagd hoe het zit met het internetgebruik onder jongeren. Met zo'n probleemstelling begint statistisch onderzoek. De probleemstelling wordt vertaald in een aantal **onderzoeksvragen**. Die vragen worden zo geformuleerd dat de antwoorden data opleveren die statistisch verwerkt kunnen worden om antwoord te geven op het gestelde probleem.

Bekijk de uitspraak hierboven van maart 2010.

- Welke onderzoeksvraag heeft het CBS zich gesteld?
- Kun je bedenken hoe het CBS dit heeft aangepakt?
- Hoe zou je zelf zo'n onderzoeksvraag aanpakken?

#### Opgave V2

Je wilt door middel van een enquête een beeld krijgen van de bestedingen van jongeren van 16-18 jaar. Formuleer minstens vijf vragen die geschikt zijn voor de enquête. Wissel jouw vragen uit met een klasgenoot en probeer elkaars vragen te verbeteren.

## Uitleg

Statistisch onderzoek wordt ingezet om informatie te krijgen over kenmerken van grote groepen. Op basis daarvan worden vaak beslissingen genomen. In Nederland doet onder andere het **Centraal Bureau voor de Statistiek (CBS)** veel statistisch onderzoek.

Een mogelijke onderzoeksvraag is: “Hoeveel jongeren van 15 tot 25 jaar bestellen toegangskaartjes via internet?”

Aan een aantal jongeren van 15 tot 25 jaar wordt de vraag gesteld:

“Heb je de afgelopen 12 maanden via internet een kaartje voor een evenement besteld?”

In 2005 antwoordde 19% van de ondervraagden ‘ja’.

In 2013 antwoordde 51% van de ondervraagden ‘ja’.

(Bron: [statline.cbs.nl](http://statline.cbs.nl), 2016)

Een onderzoeksbureau wil over een bepaalde groep informatie verzamelen. Die groep heet de populatie. Het bureau stelt een onderzoeksvraag op. Uit de populatie trekt het bureau een steekproef. Dit is het deel van de populatie aan wie de onderzoeksvraag wordt gesteld.

Alleen de informatie van de mensen uit de steekproef wordt gebruikt. Maar daarna wordt bijvoorbeeld in een interview gezegd: “51% van de Nederlandse jongeren koopt via internet kaartjes voor evenementen”.

Er wordt dus een uitspraak gedaan over de populatie, terwijl lang niet iedereen daarvan is ondervraagd.

De steekproef moet daarom goed lijken op de populatie. Alle verschillende kenmerken van de mensen in de steekproef en de populatie (zoals leeftijd, geslacht, woonplaats) moeten naar verhouding evenveel voorkomen. Dit heet een representatieve steekproef. Belangrijk is:

- Elk persoon uit de populatie moet een even grote kans hebben om in de steekproef terecht te komen; de steekproef is dan aselekt.
- De steekproef moet uit voldoende personen bestaan. Hoe groter de steekproef, hoe nauwkeuriger de uitspraken die je over de populatie kunt doen.

Bij statistisch onderzoek worden fouten gemaakt. Dat komt doordat de steekproef niet helemaal gelijk is aan de populatie. Deze fouten heten toevalsfouten.

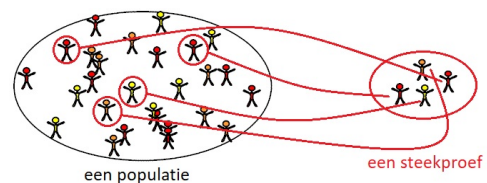
Andere fouten ontstaan door bijvoorbeeld het stellen van verkeerde enquêtevragen. Deze fouten heten systematische fouten.

Voor tips bij het opstellen van enquêtevragen, zie het **Practicum**.

### Opgave 1

Bekijk het CBS-onderzoek in de **Uitleg**.

- Hoe luidt de onderzoeksvraag?
- Hoe luidt de enquêtevraag?
- Een percentage personen bestelt dus kaartjes via internet. Waarom zal dit percentage in de populatie anders (kunnen) zijn dan in de steekproef?



Figuur 2

## Opgave 2

Leg uit of er in de omschreven situaties sprake is van een representatieve steekproef.

- Om een onderzoek te doen naar het discotheekbezoek onder 14- tot 18-jarigen kies je de leerlingen uit je klas.
- Om uit te zoeken op welke politieke partij Nederlanders stemmen bij de Tweede Kamerverkiezingen, worden uit het bevolkingsregister van Nederland willekeurig 7500 inwoners gekozen.

## Opgave 3

Leg uit of de gekozen steekproef groot genoeg is.

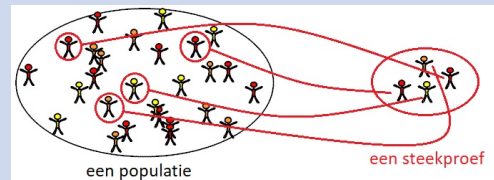
- In het buitenland komt een ziekte bij 0,01% van de mensen voor. De overheid wil weten of de ziekte ook in ons land voorkomt en onderzoekt daarom 5000 Nederlanders.
- De manager van een ijsbaan wil weten of er op woensdag meer meisjes dan jongens komen schaatsen. Hij neemt een steekproef van dertig bezoekers en telt of het meisjes of jongens zijn.

## Theorie en voorbeelden

### Om te onthouden

**Statistisch onderzoek** is onderzoek om uitspraken te doen over een bepaalde groep mensen of voorwerpen. De groep waarnaar onderzoek wordt gedaan heet **populatie**.

Een populatie is vaak heel erg groot, dus het kost te veel tijd en/of geld om iedereen of alles te onderzoeken. Er wordt dan een deel van de populatie, een **steekproef** genomen.



Figuur 3

Met de informatie uit de steekproef worden uitspraken gedaan over de hele populatie.

Het resultaat van statistisch onderzoek moet goed zijn. Daarom moet de steekproef goed lijken op de populatie. Dat is zo als

- het aantal mensen of voorwerpen in de steekproef, de **steekproefomvang** groot genoeg is: hoe groter de steekproef, hoe nauwkeuriger de resultaten zijn;
- de steekproef **aselect** is, dus elk element van de populatie een even grote kans heeft om in de steekproef te komen;
- de steekproef **representatief** is, dus alle kenmerken waar onderzoek naar wordt gedaan in de steekproef en de populatie naar verhouding even vaak voorkomen.

Bij statistisch onderzoek worden altijd **toevalsfouten** gemaakt. Dat komt doordat er een steekproef wordt genomen. Soms worden er ook andere fouten gemaakt. Dat zijn **systematische fouten**, bijvoorbeeld het stellen van verkeerde vragen of het nemen van een niet representatieve steekproef.

### Voorbeeld 1

Hoe groter de steekproef, hoe nauwkeuriger de resultaten zijn.

Een steekproef aselect nemen betekent: elk element uit de populatie moet een even grote kans hebben om in de steekproef te komen. Of een steekproef aselect is, hangt dus af van de manier waarop je de steekproef neemt.

Zijn de volgende steekproeven aselect? Leg je antwoord uit.

- Populatie: inwoners van Zoetermeer. Steekproef: willekeurig 100 bewoners kiezen uit de lijst van inwoners van Zoetermeer.
- Populatie: treinreizigers. Steekproef: vragen stellen aan willekeurige passagiers op treinroutes die willekeurig zijn gekozen.
- Populatie: ouders van leerlingen. Steekproef: een willekeurig deel van de leerlingen een vragenlijst meegeven om in te laten vullen door de ouders.

Antwoord

1. Ja. Populatie en steekproef gaan over hetzelfde en zijn willekeurig.
2. Nee. De kans dat een reiziger die tijdens de spits reist in de steekproef komt, is kleiner dan de kans dat een reiziger die buiten de spits reist in de steekproef komt.
3. Nee. De kans dat een ouder met veel kinderen op school in de steekproef komt, is groter dan de kans dat een ouder met één kind op school in de steekproef komt.

#### Opgave 4

Geef aan welke steekproef de meest betrouwbare resultaten geeft bij een populatie van 100000 mensen.

- A. De steekproef bestaat uit 3% van de populatie.
- B. Er zijn 40 enquêteurs die elk 95 mensen uit de populatie bevragen.
- C. Er komen in een periode van 10 dagen gemiddeld 350 ingevulde vragenlijsten per dag binnen.

#### Opgave 5

Zijn de volgende steekproeven aselect? Leg je antwoord uit.

- a Een onderzoek onder treinreizigers, waarbij de vragenlijsten willekeurig worden verstuurd naar bezitters van een OV-chipkaart. Ga ervan uit dat iedere reiziger maar één OV-chipkaart bezit.
- b Een onderzoek onder leerlingen van 20 scholen, waarbij van elke school 10% van de leerlingen willekeurig wordt gekozen.
- c Een onderzoek dat gebruik maakt van vragenlijsten die zijn ingevuld door willekeurig gekozen mensen van 65 jaar en ouder.
- d Een onderzoek naar de tevredenheid van abonnees van een telefoonbedrijf, door willekeurig gekozen klanten die de helpdesk bellen te ondervragen.

#### Voorbeeld 2

De kenmerken van de populatie moeten in verhouding net zo vaak voorkomen in de steekproef. Dit heet een representatieve steekproef.

Zijn de volgende steekproeven representatief? Geef ook aan of de steekproef aselect is. Een toevallige fout hoef je niet te noemen.

1. Er wordt onderzoek gedaan naar het stemgedrag voor politieke partijen van Nederlanders. Daarvoor worden aselect 8000 Nederlanders gekozen uit het bevolkingsregister.
2. Er wordt onderzoek gedaan naar de hoeveelheid verf in blikken. Van alle geproduceerde blikken wordt steeds het tiende blik gewogen.
3. Er wordt onderzoek gedaan naar het aantal uren dat ouders van de kinderen van een school van huis zijn. De ouders die op een ouderavond komen worden ondervraagd.

Antwoord

1. De steekproef is waarschijnlijk representatief. Want de steekproef is aselect en is zo groot dat alle politieke meningen in Nederland er wel zo'n beetje in voor zullen komen.
2. De steekproef is waarschijnlijk wel representatief. Want de steekproef is uit de hele populatie, de steekproefomvang is goed en meestal maakt het niet van welke blikken je de hoeveelheid meet. (Het zou natuurlijk kunnen dat elk tiende blik door één bepaalde vulkraan wordt gevuld en de rest door andere...). De steekproef is niet aselect.
3. De steekproef is niet representatief. Want waarschijnlijk komen er naar verhouding minder ouders die veel van huis zijn naar de ouderavond. De steekproef is ook niet aselect.

### Opgave 6

Geef van de volgende steekproeven aan of ze representatief en/of aselekt zijn. Leg je antwoord uit.

- Een leverancier van koffie doet onderzoek naar de tevredenheid van zijn klanten over de koffie. Hij ondervraagt alle klanten die in zijn winkel komen.
- De RDW is de organisatie die beslist welke voertuigen in Nederland worden toegelaten op de wegen. De RDW doet onderzoek naar voertuigen. De RDW neemt daarom willekeurig een grote steekproef uit de lijst van alle kentekens van voertuigen en stuurt de eigenaars van de voertuigen een vragenlijst over winterbanden.
- Een onderzoeksbureau doet onderzoek naar gebruik van sociale media onder jongeren. Het bedrijf verspreidt een enquête via sociale media.

## Verwerken

### Opgave 7

Geef van de volgende steekproeven aan of ze aselekt en/of representatief zijn. Licht je antwoord toe.

- Om een onderzoek te doen naar het theaterbezoek onder 14- tot 18-jarigen kies je de leerlingen van jouw eigen klas.
- Om de gemiddelde onderhoudskosten van een bepaald type auto te bepalen worden autobezitters via het tijdschrift 'Autokampioen' gevraagd de onderhoudskosten door te geven.

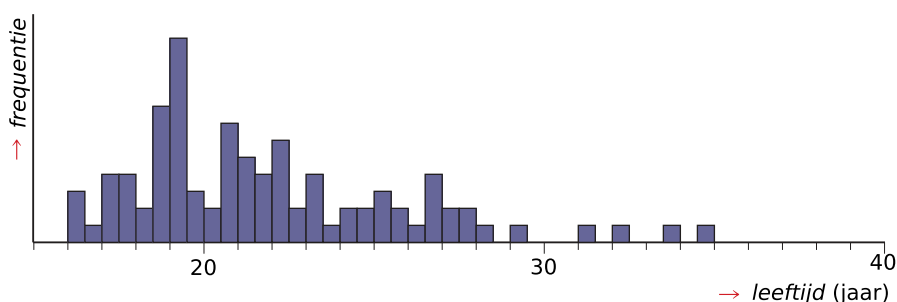
### Opgave 8

Bij het doen van metingen voor een onderzoek ontstaan soms fouten.

- Geef een voorbeeld van een systematische fout bij het meten van de temperatuur.
- Geef een voorbeeld van een toevalsfout bij het meten van de temperatuur.
- Geef twee voorbeelden van een systematische fout bij het meten van iemands gewicht in kilo's. Geef ook een voorbeeld van een toevalsfout.

### Opgave 9

Er wordt een concert gegeven met meer dan tienduizend bezoekers. Gegeven is het histogram van de frequenties van de leeftijden van de bezoekers binnen de steekproef. De steekproefomvang is 50. De steekproef is aselekt genomen.



Figuur 4

- Is de steekproef representatief? Licht je antwoord toe met twee argumenten.
- Hoe kan de steekproef wel representatief worden gemaakt?

### Opgave 10

Er wordt onderzoek gedaan naar het aantal Nederlanders dat Fries spreekt. Er wordt een steekproef genomen van 1200 deelnemers: uit elke provincie aselekt 100 inwoners.

- Lever commentaar op deze steekproefsamenstelling.
- Hoe kan het beter?

## Opgave 11

In een buurt staan rijtjeshuizen in blokken van vijf. De huizen hebben even nummers en zijn genummerd van 2-100.

De gemeente wil inzicht krijgen in het energieverbruik. Ze ondervragen daarvoor de bewoners van vijf huishoudens met de nummers 12, 20, 42, 60 en 82. Is deze steekproef representatief?

## Opgave 12

In een onderzoek naar het voedingspatroon van de Nederlandse jeugd van 16 - 22 jaar wordt een aselechte steekproef van 10000 jongeren uit die leeftijdscategorie getrokken. Al deze jongeren krijgen een schriftelijke enquête naar hun huisadres gestuurd.

- Is de steekproef representatief?
- Het percentage enquêtes dat ingevuld terug komt, heet respons. Het onderzoeksbureau krijgt 4831 enquêtes ingevuld terug. Hoe groot is de respons?
- Geef minstens twee redenen waarom er enquêtes niet zijn ingevuld (non-respons).
- Op een bepaalde vraag wordt op 94% van de ingevulde enquêtes met 'ja' geantwoord. Mag er nu worden geconcludeerd dat 94% van de jongeren van 16 - 22 jaar het daarmee eens is?
- Een van de conclusies van het onderzoek is: de Nederlandse jeugd drinkt de eerste keer wodka op een gemiddelde leeftijd van 19,3 jaar. Lever commentaar.

## Toepassen

### Opgave 13: Risico op hartinfarct

In een onderzoek wordt de volgende conclusie getrokken:

"Het risico op een hartinfarct van de gemiddelde Amerikaanse man wordt met ongeveer 45% verlaagd door het slikken van aspirine."

Van het onderzoek is het volgende bekend.

In de jaren 1982 - 1988 werd onder 22000 mannelijke Amerikaanse artsen onderzoek gedaan naar de invloed van aspirine op hart- en vaatziekten. De helft van de mannen gebruikte om de dag 300 mg aspirine, wat ongeveer gelijk staat aan een "gewoon" aspirientje. De andere helft slikte een placebo ('fopmiddel').

Van de aspirineslikkers kregen 104 personen een hartinfarct, van de placeboslikkers waren dat er 189.

Dat dit grote verschil aan toevalsfouten was te wijten vond men uitgesloten vanwege het grote aantal mensen dat aan de studie meewerkte.

- Waarom mag de conclusie niet worden getrokken uit het onderzoek?
- Waarom wordt er van placebo's gebruik gemaakt?
- Hoeveel procent van de 11000 aspirineslikkers heeft waarschijnlijk baat gehad bij het slikken van aspirine?
- Volgens de tekst hierboven wordt de kans op een hartinfarct met 45% verlaagd. Klopt dat?
- Waarom zal de conclusie niet direct zijn: 'Adviseer de populatie aspirine te gaan slikken?'

## Testen

### Opgave 14

Bij het houden van een enquête wordt een steekproef uit een bepaalde populatie getrokken. Elke manier waarop je dit doet heeft zijn voor en tegen, vaak afhankelijk van de populatie.

- Een onderzoeksbureau onderzoekt of mensen op een krant zijn geabonneerd door een telefonische enquête waarbij de nummers aselekt uit een telefoonboek worden getrokken. Zal de uitslag van deze enquête een goed beeld geven of vindt er vertekening plaats? Motiveer je antwoord.

- b** Je houdt een telefonische enquête onder alleen mobiele nummers met 'random digit dialling'. Is dit voor elke populatie geschikt?
- c** Je houdt een enquête via een jongerenwebsite als het gaat om het uitgavenpatroon van jongeren. Goede aanpak?
- d** Je houdt een enquête via de ANWB-website als het gaat om 'rekening rijden'. Goede aanpak?
- e** Je houdt een enquête via e-mail onder de leden van een bepaalde club over contributieverhoging. Goede aanpak?
- f** Je houdt een schriftelijke enquête onder 5000 aselekt gekozen adressen. Welk nadeel heeft deze aanpak?

## Practicum

Hoe je te werk kunt gaan bij **het verzamelen van data middels een enquête** in Google-Drive zie je in het practicum:

- **Data verzamelen met een enquête**

Het **trekken van een aselekte steekproef** kun je simuleren met deze VUstat-app. Daarin kun je zien welk percentage een aselekte steekproef van bijvoorbeeld 200 (in te stellen bij 'omvang') uit een populatie van 5000 oplevert als je het percentage in de steekproef weet (instellen bij "percentage Groen"). Je kunt ook heel veel van die steekproeven laten doen en zien hoezeer de percentages gespreid liggen. Ga naar de app:



**Figuur 5**

- **Steekproeven uit ja/nee populatie**



© 2021

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt.

Email: [f.spijkers@math4all.nl](mailto:f.spijkers@math4all.nl)

Met de Math4All maatwerkdienst kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij [a.f.otten@xs4all.nl](mailto:a.f.otten@xs4all.nl) een gratis inlog voor de maatwerkdienst aanvragen.

---