

## 3.3 Populatieproportie schatten

### Inleiding

Wat zegt de steekproefproportie over de populatieproportie? Kun je zo maar zeggen dat als je 10% rotte sinaasappels aantreft in jouw steekproef, dat dan ook 10% van alle sinaasappels rot is? Met welke betrouwbaarheid kan dat wel? Hoe bereken je de foutmarge?

Over deze vragen gaat dit onderdeel.



Figuur 1

### Je leert in dit onderwerp

- vanuit een steekproefproportie een bijbehorende standaardafwijking berekenen;
- een betrouwbaarheidsinterval van een populatieproportie berekenen met behulp van de vuistregels.

### Voorkennis

- de begrippen onderzoek, steekproef, populatie en representatief;
- de vuistregels van de normale verdeling en van elke steekproevenverdeling.

### Verkennen

#### Opgave V1

Je wilt bepalen hoeveel procent van de Nederlandse vrouwen tussen 15 en 25 jaar rookt. In een aselechte steekproef van 1200 vind je 352 vrouwen die roken.

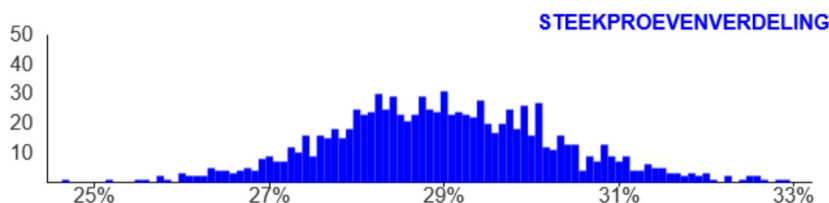
- a** Laat zien dat de steekproefproportie  $p$  gelijk aan 29,3% is.

In een tweede aselechte steekproef van 1200 vind je 338 vrouwen die roken.

- b** Bereken ook nu de steekproefproportie  $p$ .

Je ziet dat het steekproefresultaat afhankelijk is van toeval. Stel dat je in totaal 1000 van deze steekproeven neemt en je maakt dit staafdiagram van alle steekproefproporties.

Neem aan dat de steekproefproporties normaal verdeeld zijn met gemiddelde 0,29 en standaardafwijking 0,013.



Figuur 2

- c** Tussen welke waarden liggen dan de steekproefproporties van 95% van de steekproeven?
- d** Kun je nu met een betrouwbaarheid van 95% een uitspraak doen over het percentage vrouwen tussen 15 en 25 jaar die roken?
- e** Waarom zullen er in de praktijk nooit 1000 van die steekproeven worden genomen?

In de populatie vrouwen rookt een bepaald percentage, je weet het alleen niet.

Maar in het **Practicum** kun je bij een populatieproportie die je zelf instelt, nagaan hoe de steekproefproporties zich gedragen. Je gebruikt VUstat.

- f Voer zo'n simulatie van 1000 steekproeven uit een populatie van 1200 uit met een populatieproportie van 0,30. Ga na dat de steekproefproporties ook rond de  $0,30 \cdot 1200 = 360$  uitkomen.

## Uitleg

Stel dat men wil onderzoeken hoeveel procent van de populatie jongeren vanaf 12 t/m 16 jaar lid is van een sportvereniging (geen fitnessclub). Dit heet de populatieproportie.

Het is ondoenlijk om de hele populatie te bevragen. Dus wordt er een (aselecte, representatieve) steekproef getrokken van  $n = 500$  jongeren van 12 t/m 16. Er blijken 360 jongeren in de steekproef "ja" te antwoorden op de vraag of ze lid zijn van een sportvereniging.

De steekproefproportie is  $p = \frac{360}{500} = 0,72$  ofwel 72%,

Maar wat zegt dit over de populatieproportie?

Bij **Verkennen V1** heb je kunnen zien dat het trekken van veel even grote steekproeven uit een populatie een verzameling steekproefproporties  $p$  oplevert die normaal is verdeeld met een gemiddelde in de buurt van de populatieproportie. In de bijbehorende standaardafwijking gebruik je  $p$  in plaats van de onbekende populatieproportie:

$$\sigma = \sqrt{\frac{p \cdot (1-p)}{n}}$$

Volgens de vuistregels van een normale verdeling ligt in 95% van de steekproeven  $p$  minder dan  $2\sigma$  van de populatieproportie af. Nu is die populatieproportie nog onbekend, je hebt alleen  $p$ .

En als  $p$  minder dan  $2\sigma$  van de populatieproportie af ligt, ligt de populatieproportie ook minder dan  $2\sigma$  van  $p$  af. Dat gaat goed in 95% van de gevallen.

In de beschreven situatie is  $\sigma = \sqrt{\frac{0,72 \cdot (1-0,72)}{500}} \approx 0,020$ .

Dus de populatieproportie ligt dan (met een betrouwbaarheid van 95%) minder dan  $2\sigma = 2 \cdot 0,020 = 0,04$  van  $p = 0,72$  af, dus tussen  $0,72 - 0,04 = 0,68$  en  $0,72 + 0,04 = 0,76$ .

Of in procenten tussen 68% en 76%.

Dit heet het 95% betrouwbaarheidsinterval voor de populatieproportie.

### Opgave 1

Bekijk **Uitleg**. Ga uit van het beschreven onderzoek onder jongeren vanaf 12 t/m 16 jaar.

- Reken zelf  $p$  en de standaardafwijking  $\sigma$  van de steekproevenverdeling na.
- Gebruik in het **Practicum** de simulatie van VUstat. Stel een populatieproportie van 72% in en simuleer eerst één steekproef met  $n = 500$  en vervolgens een grote steekproevenverdeling. Ga na, dat het gemiddelde van de steekproevenverdeling ongeveer op 72% uitkomt en dat de standaardafwijking van de steekproevenverdeling voldoet aan de formule in de uitleg.
- Natuurlijk weet je in de praktijk niet hoe groot de populatieproportie is, dus kun je zo'n simulatie dan niet gebruiken.

Maar ga na dat je door nieuwe simulaties te doen met populatieproportie 72% en  $n = 500$  je wel steeds op ongeveer dezelfde gemiddelde steekproevenproportie en standaardafwijking uitkomt.

### Opgave 2

Er wordt een onderzoek gedaan naar welk deel van festivalgangers ouder dan 40 jaar is. Bij een festival is door middel van een steekproef aan 150 bezoekers de leeftijd gevraagd. Van deze groep blijken 50 bezoekers 40 jaar of ouder te zijn. Bereken in vier decimalen nauwkeurig de bijbehorende steekproefproportie en de standaardafwijking van de steekproevenverdeling. Geef ook het 95% betrouwbaarheidsinterval voor de populatieproportie.

### Opgave 3

Er wordt in **Uitleg** geredeneerd met een betrouwbaarheid van 95%.

- a Tussen welke grenzen ligt in 68% van de steekproeven het 'Ja'-deel van de populatie?
- b Tussen welke grenzen ligt in bijna 100% van de steekproeven het 'Ja'-deel van de populatie?

### Opgave 4

Om te bepalen hoeveel procent van de Nederlanders linkshandig is wordt een aselechte steekproef van 1500 Nederlanders getrokken. Daarvan waren er 136 linkshandig.

- a Bereken het deel van de steekproef dat linkshandig is.
- b Bereken de standaardafwijking van steekproevenverdeling in de steekproef. Rond je antwoord af op vier decimalen.
- c Tussen welke waarden ligt met 95% betrouwbaarheid het deel van de populatie dat linkshandig is? Geef je antwoord in procenten in één decimaal nauwkeurig.

## Theorie en voorbeelden

### Om te onthouden

Bij statistisch onderzoek worden uitspraken gedaan over de populatie.

Er wordt vaak gebruik gemaakt van vragen met maar 2 mogelijke antwoorden. Het deel van de steekproef dat dan het ene antwoord geeft heet de **steekproefproportie**,  $p$ . Andere steekproeven van grootte  $n$  zullen vaak andere waarden van  $p$  geven. Maar deze steekproefproporties zijn altijd normaal verdeeld. Deze normale verdeling heet **steekproevenverdeling** met standaardafwijking  $\sigma$ .

Er moet een conclusie worden getrokken over het deel van de populatie dat dit antwoord zou geven, de **populatieproportie**. De standaardafwijking van de steekproevenverdeling is:

$$\sigma = \sqrt{\frac{p \cdot (1-p)}{n}}$$

Voor de populatieproportie geldt dan:

- In 68% van de steekproeven ligt de populatieproportie tussen  $p - \sigma$  en  $p + \sigma$ . Dit heet het 68%-**betrouwbaarheidsinterval** van  $p$  en de 68% heet de **betrouwbaarheid**.
- In 95% van de steekproeven ligt de populatieproportie tussen  $p - 2 \cdot \sigma$  en  $p + 2 \cdot \sigma$ . Dit heet het 95%-betrouwbaarheidsinterval van  $p$ .
- In bijna 100% van de steekproeven ligt de populatieproportie tussen  $p - 3 \cdot \sigma$  en  $p + 3 \cdot \sigma$ .

De formule voor  $\sigma$  en de vuistregel voor een betrouwbaarheid van 95% vind je op deze **formulekaart**. In dit geval is  $2 \cdot \sigma$  de **foutmarge**.

### Voorbeeld 1

Er wordt onderzoek gedaan naar het percentage voorstanders van de hypotheekrenteaftrek onder stemgerechtigden. Gekozen is een betrouwbaarheid van 95%. Er wordt een steekproef genomen met een omvang van 1000. Het aantal voorstanders 570. Bereken het betrouwbaarheidsinterval van de populatieproportie van de voorstanders. Leg uit wat dit betekent.

Antwoord

De steekproefproportie is  $p = \frac{570}{1000} = 0,57$ .

$$\sigma = \sqrt{\frac{0,57 \cdot (1-0,57)}{1000}} \approx 0,0157$$

Het 95%-betrouwbaarheidsinterval ligt tussen  $p - 2 \cdot \sigma$  en  $p + 2 \cdot \sigma$ .

Dus tussen  $0,57 - 2 \cdot 0,0157 \approx 0,539$  en  $0,57 + 2 \cdot 0,0157 \approx 0,601$ .

Je noemt die  $2 \cdot 0,0157 \approx 3,1\%$  wel de foutmarge van de schatting van de populatieproportie.

De betekenis hiervan kun je op meerdere manieren omschrijven. Bijvoorbeeld:

- Het percentage voorstanders van hypotheekrenteaf trek wordt geschat tussen de 53,9 en 60,1.
- Bij 100 goede steekproeven, zal de populatieproportie ongeveer 95 keer tussen de zo berekende grenzen liggen.

### Opgave 5

Bij een onderzoek naar de slagingskans voor het rijexamen wordt in een bepaald jaar van een steekproef van 800 pogingen vastgesteld of het examen is gehaald of niet. Van die 800 pogingen bleken er 683 succesvol te zijn.

- Bereken de steekproefproportie. Geef je antwoord exact en afgerond op twee decimalen.
- Bereken de standaardafwijking van de steekproevenverdeling. Rond je antwoord af op vier decimalen.
- Bepaal de foutenmarge bij een betrouwbaarheid van 95%.

### Opgave 6

De Consumentenbond wil weten of een bepaald type laptop minstens acht uur op de accu kan werken. De bond test 50 aselekt getrokken laptops van dat type. Het blijkt dat 41 van die laptops inderdaad minstens acht uur werkt op de accu.

- Hoeveel bedraagt de proportie laptops  $p$  in de steekproef, dat acht uur lang op de batterij werkt?
- Bereken de steekproefstandaardafwijking  $\sigma$  van de steekproefproportie volgens de formule in de uitleg. Rond je antwoord af op vier decimalen.
- Hoe groot is het 95%-betrouwbaarheidsinterval van de steekproefproportie?
- Schrijf nu de uitspraak op die je met 95% betrouwbaarheid kunt doen.

### Voorbeeld 2

Een fabrikant laat statistisch onderzoek doen. Hij wil weten hoeveel procent van de tennisballen die hij laat maken, zwaarder is dan 59,4 gram. Uit een kleine steekproef komt dat ongeveer 6% van de tennisballen te zwaar is. De fabrikant neemt vervolgens een grotere steekproef om met een betrouwbaarheid van 95% te kunnen vaststellen dat tussen 5% en 7% van zijn tennisballen zwaarder is dan 59,4 gram.

Hoe groot moet de steekproefomvang van de tweede steekproef zijn?

Antwoord

Ook voor de grotere steekproef moet dan  $p = \frac{0,05+0,07}{2} = 0,06$ , dezelfde 6% als voor de kleinere steekproef. Omdat het 95%-betrouwbaarheidsinterval ligt tussen  $0,06 - 2\sigma$  en  $0,06 + 2\sigma$  is  $\sigma = \frac{1}{2} \cdot 0,01 = 0,005$ .

Dit betekent:  $0,005 = \sqrt{\frac{0,06 \cdot (1-0,06)}{n}}$ .

Oplossen geeft:  $n = 2256$ .

Dus er moeten minstens 2256 tennisballen worden getest.

### Opgave 7

Bekijk [Voorbeeld 2](#).

- Waar komt  $\sigma = \frac{1}{2} \cdot 0,01$  vandaan?
- Bereken zelf de waarde van  $n$ .

### Opgave 8

Bekijk **Voorbeeld 2**. Als de fabrikant het 95% betrouwbaarheidsinterval tussen 4% en 8% had willen hebben, hadden er dan meer of minder ballen moeten worden getest? Licht je antwoord toe.

## Verwerken

### Opgave 9

Bij een aselechte steekproef in jouw provincie blijken onder 1500 geënquêteerden er 833 tegen de aanleg van een provinciale weg te zijn.

- Bereken de steekproefproportie.
- Bereken de standaardafwijking van deze steekproevenverdeling. Rond je antwoord af op vier decimalen.

### Opgave 10

Bij een statistisch onderzoek is een steekproefproportie  $p \approx 0,40$  gevonden en de standaardafwijking van de steekproevenverdeling  $\sigma \approx 0,025$ . Hiermee kan de populatieproportie met een zekere betrouwbaarheid worden geschat.

- Welke grenzen heeft het 68%-betrouwbaarheidsinterval?
- Welke grenzen heeft het 95%-betrouwbaarheidsinterval?
- Welke grenzen heeft het bijna 100%-betrouwbaarheidsinterval?

### Opgave 11

Bij een aslechte steekproef op een grote school blijken in een steekproef van 150 leerlingen er 31 bloedgroep O te hebben.

- Bepaal het 95% betrouwbaarheidsinterval voor de populatieproportie.
- Wat zijn de gevolgen voor de breedte van het betrouwbaarheidsinterval als blijkt dat het aantal leerlingen met bloedgroep O niet 31 maar 32 is?
- Wat zijn de gevolgen voor de grenzen van het 95% betrouwbaarheidsinterval als blijkt dat het aantal leerlingen met bloedgroep O niet 31 maar 32 is?

### Opgave 12

Er worden 500 aselekt getrokken laptops onderzocht. Er blijken 406 van die laptops minstens 8 uur te werken op de batterij. Welke uitspraak kun je nu met een betrouwbaarheid van 95% doen?

### Opgave 13

Uit een enquête in opdracht van de Stichting tegen Kanker van maart/april 2007 onder 1988 Belgen bleek 61% voorstander te zijn van het rookvrij maken van cafés. In oktober 2006 was dat nog 55% van de even grote groep van toen ondervraagde personen.

- Bepaal bij het onderzoek van oktober 2006 het betrouwbaarheidsinterval bij een betrouwbaarheid van 95%. Geef je antwoord in procenten.
- Bepaal bij het onderzoek van maart/april 2007 het betrouwbaarheidsinterval bij een betrouwbaarheid van 95%. Geef je antwoord in procenten.
- Welke conclusie kun je nu trekken?

### Opgave 14

- Bij een steekproef van 500 lampen vond de controledienst 56 kapotte lampen. Wat kun je met 95% betrouwbaarheid zeggen over het percentage kapotte lampen van de hele populatie?
- Bij een steekproef van 500 lampen vond de controledienst 444 goede lampen. Wat kun je met 95% betrouwbaarheid zeggen over het percentage goede lampen van de hele populatie?
- Vergelijk de antwoorden van de vragen a en b met elkaar. Wat valt je op? Geef een verklaring.

## Toepassen

### Opgave 15: Opiniepeilingen

Voorafgaande aan verkiezingen worden opiniepeilingen gehouden. Daarbij worden door een onderzoeksbureau 2000 aselekt getrokken Nederlanders gevraagd naar de partij van hun voorkeur. Een partij gaat in zo'n opiniepeiling van 30 naar 31 zetels (van de 150 zetels).

Gebruik een betrouwbaarheidsniveau van 90%. Daar hoort een foutmarge van  $1,65 \cdot \sigma$  bij.

Onderzoek of er reden tot blijdschap voor deze partij is vanwege deze peiling.

### Opgave 16: Hoe groot moet je steekproef zijn?

Je wilt een populatieproportie bepalen met een steekproef. Je wilt dat de foutmarge maximaal 0,03 is bij een 95% betrouwbaarheid. Ga uit van  $p_{\text{steekproef}} = 0,5$  en een stad met 50000 inwoners. Hoeveel mensen moet je minstens ondervragen?

## Testen

### Opgave 17

Helmond is een stad van ongeveer 90000 inwoners en de stad heeft een stadspanel met 1300 deelnemers. Op een verzoek van de gemeente om een vragenlijst over de lokale omroep in te vullen hebben 839 mensen gereageerd.

Uit deze meting blijkt dat 37% de lokale omroep niet kent.

- Welke uitspraak kun je doen met 95%-betrouwbaarheid over de onbekendheid van de lokale omroep in Helmond?
- Welke uitspraak kun je met 95% betrouwbaarheid doen over de bekendheid van de lokale omroep in Helmond?
- Welke kanttekeningen kun je bij dit onderzoek maken?

### Opgave 18

Bij een marktonderzoek wordt gekeken naar de belangstelling voor elektrische auto's onder particulieren in Nederland. In een aselechte steekproef worden 1660 mensen benaderd en hiervan zeggen 917 particulieren dat ze de overstap naar een elektrische auto serieus overwegen.

De onderzoekers willen het percentage belangstellenden met een betrouwbaarheid van 95% vaststellen met een marge van 0,5%.

Hoe groot moet hun steekproef dan zijn?

## Practicum

Een applet van een groot aantal steekproeven (steekproefgrootte in te stellen) uit een populatie waarvan de populatieproportie is in te stellen. Het doel is om te laten zien dat die steekproefproporties normaal verdeeld liggen. Zo kun je een **populatieproportie schatten**, met een bijbehorend een betrouwbaarheidsinterval en een foutenmarge.

- [Simulatie steekproeven om een populatieproportie te schatten](#)

Dit practicum is ontwikkeld door Piet van Blokland en Carel van de Giessen, zie [www.vusoft.eu](http://www.vusoft.eu)

### Simulaties van schattingen in Excel

Het trekken van aselechte steekproeven uit een populatie is ook te bekijken via het practicum:

- [Steekproeven en uitspraken](#)




Figuur 3



© 2024

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. Klik op  in de marge bij de betreffende opgave. Uw mailprogramma wordt dan geopend waarbij het emailadres en onderwerp al zijn ingevuld. U hoeft alleen uw opmerkingen nog maar in te voeren.

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

Met de Math4All Foliostaat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij [a.f.otten@math4all.nl](mailto:a.f.otten@math4all.nl) een gratis inlog voor de maatwerkdienst aanvragen.

---

