3.3 Populatieproportie

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

- de standaardafwijking (met een formule) van de steekproevenverdeling van de steekproefproportie uitrekenen;
- bij een gegeven betrouwbaarheid de foutmarge van de schatting van de steekproefproportie berekenen;
- de grenzen van een 90%, 95% en 99% betrouwbaarheidsinterval van een schatting van de populatieproportie berekenen.

Voorkennis

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

Verkennen

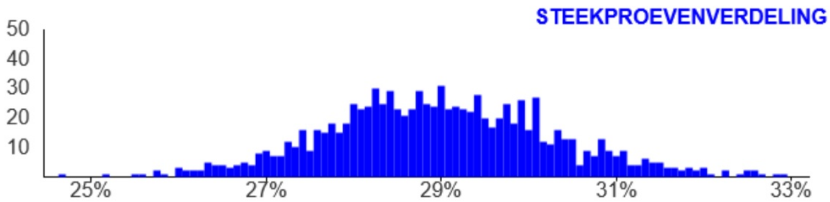
Opgave V1

Je wilt bepalen hoeveel procent van de Nederlandse vrouwen tussen 15 en 25 jaar rookt. In een aselecte steekproef van 1200 vind je 348 vrouwen die roken.

a Laat zien dat de steekproefproportie $p_{\text{steekproef}}$ gelijk aan 29% is.

b Je hebt alleen maar deze ene steekproef en je weet dat het steekproefresultaat afhankelijk is van toeval. Simuleer daarom een steekproevenverdeling: neem 1000 keer een steekproef van 1200 en teken het staafdiagram van de steekproefproporties en de bijpassende normaalkromme.

c Een andere simulatie heeft deze steekproevenverdeling opgeleverd. Neem aan dat de steekproefproporties normaal verdeeld zijn met gemiddelde 0,29 en standaardafwijking 0,013. Tussen welke waarden liggen dan de steekproefproporties van 95% van de vrouwen?



Figuur 2

d Geef het betrouwbaarheidsinterval van de populatieproportie bij betrouwbaarheidsniveau van 95%.

e Doe nu met een betrouwbaarheid van 95% een uitspraak over het percentage vrouwen tussen 15 en 25 jaar die roken.

Uitleg 1

Bij statistische onderzoeken komen vaak vragen voor met maar twee mogelijke antwoorden, bijvoorbeeld:

- Ben je man of vrouw?
- Ben je ouder dan 40 jaar?
- Ben je getrouwd?

Als 39% van de ondervraagden op de vraag "Ben je man?" met "ja" antwoordt, geldt: Het deel van de ondervraagden dat man is, is $p_{\text{steekproef man}} = 0{,}39$. Dit heet de steekproefproportie mannen.

Bij "nee" hoort in dit geval dus $p_{\text{steekproef vrouw}} = 0{,}61$.

Deze twee waarden van p zijn bij elkaar altijd 1.

Elke keer dat er een steekproef wordt genomen, kan het deel van de mensen dat 'ja' antwoordt, een iets andere waarde hebben. Deze waarden van p vormen een steekproevenverdeling en zijn normaal verdeeld.

De steekproevenverdeling heeft een standaardafwijking σ die je kunt berekenen uit $p_{\text{steekproef}}$ en het aantal mensen n per steekproef:

$$\sigma = \sqrt{\frac{p_{\text{steekproef}} \cdot (1 - p_{\text{steekproef}})}{n}}$$

Deze formule staat op een **formulekaart**.

Stel een groot warenhuis in Den Haag onderzoekt zijn populatie klanten. Aan 50 mensen wordt de vraag gesteld: "Woont u in Den Haag?" waarop 36 ondervraagden antwoorden "Ja".

Dan is $p_{\text{steekproef}} = \frac{36}{50} = 0{,}72$.

Hierbij hoort een steekproevenverdeling met een standaardafwijking van $\sigma = \sqrt{\frac{0{,}72 \cdot (1 - 0{,}72)}{50}} \approx 0{,}063$.

Opgave 1

Bekijk **Uitleg 1**. Bereken in drie decimalen nauwkeurig de standaardafwijking van de steekproevenverdeling met de volgende gegevens.

a $n = 100$ en $p_{\text{steekproef}} = 0{,}3$.

b Een steekproef van 200 personen, waarvan 130 personen 'Ja' antwoorden.

c Een populatie van 11000 personen, waarvan 12% als steekproef wordt genomen. 63% van de mensen uit de steekproef zegt '1 of meer kinderen te hebben'.

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.

Uitleg 2

Je wilt een populatie onderzoeken door het nemen van een steekproef van 50 mensen. Die mensen stel je een vraag met twee mogelijke antwoorden. Het deel $p_{\text{steekproef}}$ van de mensen uit de steekproef dat 'ja' antwoordde was 0,72, dit heet de steekproefproportie. Maar dit kan een toevallige uitkomst zijn, andere steekproeven van 50 kunnen andere waarden opleveren. De bijbehorende steekproevenverdeling heeft een standaardafwijking

$$\sigma = \sqrt{\frac{0{,}72 \cdot (1 - 0{,}72)}{50}} \approx 0{,}063.$$

Je weet nu de proportie $p_{\text{steekproef}} = 0{,}72$ en de standaardafwijking van de steekproevenverdeling. Wat kun je daarmee over de populatie zeggen?

Uit de vuistregels van de normale verdeling volgt dat 95% van de proporties in de steekproevenverdeling tussen $0{,}72 - 2 \cdot 0{,}063 \approx 0{,}59$ en $0{,}72 + 2 \cdot 0{,}063 \approx 0{,}85$ zal liggen.

Dit betekent dat in 95% van de steekproeven het deel van de populatie dat 'ja' zegt tussen 0,59 en 0,85 ligt. Dit heet het 95%-betrouwbaarheidsinterval van de populatieproportie.

Opgave 3

Bij een steekproef van 150 mensen is gebleken dat 30 van deze mensen dagelijks de wekker vroeger dan 7 uur 's ochtends zetten.

Tussen welke waarden ligt het deel van de mensen dat de wekker vroeger dan 7 uur 's ochtends zet in 95% van dergelijke steekproeven? Geef je antwoord in drie decimalen.

Opgave 4

Er wordt in **Uitleg 2** 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 5

Om te bepalen hoeveel procent van de Nederlanders linkshandig is wordt een aselecte 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 in 95% van dergelijke steekproeven het deel van de populatie dat linkshandig is? Geef je antwoord in drie decimalen.

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**, $p_{\text{steekproef}}$. Andere steekproeven van grootte n zullen vaak andere waarden van $p_{\text{steekproef}}$ geven. Maar deze steekproefproporties zijn altijd normaal verdeeld. Deze normale verdeling heet **steekproevenverdeling** met standaardafwijking σ.

Er moet een conclusie worden getrokken over het deel van de populatie dat dit antwoord zou geven, de **populatieproportie**, p. Je neemt aan dat

$p \approx p_{\text{steekproef}}$

met als bijbehorende standaardafwijking:

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

Uit de vuistregels voor de normale verdeling volgt nu:

- 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 σ 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

$p_{\text{steekproef}} = \frac{570}{1000} = 0{,}57 = p$

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

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 $0{,}57 + 2 \cdot 0{,}0157 \approx 0{,}601$.

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

- Het percentage voorstanders van hypotheekrenteaftrek ligt in 95% van deze steekproeven tussen de 53,9 en 60,1.
- In 95% van de steekproeven is $57 \pm 3{,}1$% van de populatie voorstander van hypotheekrenteaftrek. Je noemt die 3,1% wel de foutmarge van de schatting van de populatieproportie.
- Bij 100 goede steekproeven, zal het percentage voorstanders in de steekproef ongeveer 95 keer tussen 53,9 en 60,1 liggen.

Opgave 6

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.

a Bereken de steekproefproportie. Geef je antwoord exact en afgerond op twee decimalen.

b Bereken de standaardafwijking van de steekproevenverdeling. Rond je antwoord af op vier decimalen.

c Bepaal de foutenmarge bij een betrouwbaarheid van 95%.

Opgave 7

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

a Hoeveel bedraagt de proportie laptops $p_{\text{steekproef}}$ in de steekproef, dat acht uur lang op de batterij werkt?

b Bereken de steekproefstandaardafwijking σ van de steekproefproportie volgens de formule in de uitleg. Rond je antwoord af op vier decimalen.

c Hoe groot is het 95%-betrouwbaarheidsinterval van de steekproefproportie?

d 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 $p_{\text{steekproef}} = p = \frac{0,05+0,07}{2} = 0,06$, 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 8

Bekijk **Voorbeeld 2**.

a Waar komt $\sigma = \frac{1}{2} \cdot 0,01$ vandaan?

b Bereken zelf de waarde van n.

Opgave 9

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 10

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

a Bereken de steekproefproportie.

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

Opgave 11

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.

a Welke grenzen heeft het 68%-betrouwbaarheidsinterval?

b Welke grenzen heeft het 95%-betrouwbaarheidsinterval?

c Welke grenzen heeft het bijna 100%-betrouwbaarheidsinterval?

Opgave 12

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

a Bepaal het 95% betrouwbaarheidsinterval voor de populatieproportie.

b Wat zijn de gevolgen voor de breedte van het betrouwbaarheidsinterval als blijkt dat het aantal leerlingen met bloedgroep O niet 31 maar 32 is?

c 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 13

Er worden 500 aselect 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 14

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.

a Bepaal bij het onderzoek van oktober 2006 het betrouwbaarheidsinterval bij een betrouwbaarheid van 95%. Geef je antwoord in procenten.

b Bepaal bij het onderzoek van maart/april 2007 het betrouwbaarheidsinterval bij een betrouwbaarheid van 95%. Geef je antwoord in procenten.

c Welke conclusie kun je nu trekken?

Opgave 15

a 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?

b 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?

c Vergelijk de antwoorden van de vragen a en b met elkaar. Wat valt je op? Geef een verklaring.

Toepassen

Opgave 16: Opiniepeilingen

Voorafgaande aan verkiezingen worden opiniepeilingen gehouden. Daarbij worden door een onderzoeksbureau 2000 aselect 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 17: 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 500.000 inwoners. Hoeveel mensen moet je minstens ondervragen?

Testen

Opgave 18

Helmond is een stad van ongeveer 90000 inwoners en de stad heeft een stadspanel met 1300 deelnemers. Op een verzoek van de gemeente een vragenlijst in 2012 over de lokale omroep in te vullen hebben 839 mensen gereageerd.
Uit deze meting blijkt dat 37% de lokale omroep niet kent.

a Welke uitspraak kun je doen met 95%-betrouwbaarheid over de onbekendheid van de lokale omroep in Helmond?

b Welke uitspraak kun je met 95% betrouwbaarheid doen over de bekendheid van de lokale omroep in Helmond?

c Welke kanttekeningen kun je bij dit onderzoek maken?

Opgave 19

Bij een marktonderzoek wordt gekeken naar de belangstelling voor elektrische auto's onder particulieren in Nederland. In een aselecte 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

Figuur 3

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**

Simulaties van schattingen in Excel

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

- **Steekproeven en uitspraken**

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`

Met de Math4All maatwerkdienst kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij `a.f.otten@xs4all.nl` een gratis inlog voor de maatwerkdienst aanvragen.