



Hoofdstuk 1

Betrouwbare uitspraken op basis van gegevens

1.1 Betrouwbare gegevens verzamelen

- 1.1.1 Basisbegrippen
- 1.1.2 Het samenstellen van een steekproef: steekproefontwerp

1.2 Steekproefvariabiliteit en steekproefverdeling

1.3 De betrouwbaarheid van een steekproefresultaat cijfermatig uitdrukken

- 1.3.1 Foutenmarge en betrouwbaarheidsinterval
- 1.3.2 Betrouwbaarheidsintervallen bepalen met een tabel



Opdracht 1 bladzijde 9

'De enquête "Energiegebruik in het Vlaamse huishouden" werd (...) uitgevoerd om een goed inzicht te verkrijgen in het energiebewustzijn en -gedrag van de Vlaamse huishoudens. (...) Medio 1998 werden 1000 Vlaamse huishoudens geënuquêteerd. In 50 gemeenten, verspreid over de 5 Vlaamse provincies werden telkens 20 huis-aan-huis enquêtes afgenomen. (...) In verband met het warmwaterverbruik werd vastgesteld dat in het gemiddelde Vlaamse huisgezin wekelijks 7,1 douches en 4,2 baden worden genomen.'

Geef bij de bovenstaande uitspraak de beoogde populatie, de variabele die wordt onderzocht, de steekproef, inclusief de steekproefgrootte.

Populatie: alle Vlaamse huishoudens.

Onderzochte variabelen: het wekelijkse aantal douches en baden.

Steekproef: 1000 Vlaamse huishoudens (de onderzochte eenheden zijn dus huishoudens, niet de personen die de enquêtes invulden).

Opdracht 2 bladzijde 10**Peilingen in een winkelstraat**

Je werkt als jobstudent voor een firma die opiniepeilingen uitvoert in opdracht van haar klanten. Voor een bepaalde peiling ga je enquêtes afnemen in een drukke winkelstraat. De populatie waarover een uitspraak moet worden gedaan, is 'mensen tussen 15 en 75'.

- 1 Zijn 'winkelstraatbezoekers' altijd representatief voor de beoogde populatie? Indien niet, welke groepen zijn over- of ondervertegenwoordigd? Welke factoren beïnvloeden de representativiteit?

Winkelstraatbezoekers vormen geen representatieve deelgroep van de onderzochte populatie: niet elk element uit de populatie heeft dezelfde kans om in de steekproef voor te komen.

De samenstelling van de steekproef zal o.a. afhankelijk zijn van de dag en het uur van enquêteren. Tijdens de week overdag vind je er nauwelijks mensen die fulltime werken, maar eerder mensen die deeltijds werken of zonder job, mensen met pensioen, jongeren ... In het algemeen heb je er minder kans om er mensen aan te treffen die ziek zijn of een handicap hebben. Mensen die armer zijn vind je er mogelijk ook minder, afhankelijk van de buurt waar de winkelstraat zich bevindt.

- 2 Zijn er mensen die je liever niet of wel zou aanspreken om een enquête af te nemen?

Eenzijds kunnen persoonskenmerken een invloed hebben: leeftijd, geslacht, huidskleur, voorkomen, gedrag ... Mensen alleen zijn bovendien gemakkelijker aan te spreken, dan mensen in groep.

Opdracht 3 bladzijde 10**Peilingen via een website**

Een milieuvereniging heeft op haar website een enquête staan die iedereen kan invullen. Nadat er een duizendtal enquêtes werden ingevuld, schrijft de vereniging in een persartikel: '60 % van de Vlamingen vindt dat er meer aandacht naar milieu moet gaan.'

Denk je dat deze steekproef betrouwbaar is? Waarom (niet)?

Niet iedereen heeft internet. Wellicht worden sites van milieuorganisaties vooral bezocht door mensen die om het milieu bekommerd zijn en van die groep zullen alleen de meer gemotiveerde bezoekers de moeite doen om een enquête in te vullen. Al deze groepen vragen wellicht om meer aandacht voor het milieu.

Wie niet met de milieuproblematiek bezig is, wie industrie en concurrentiekracht belangrijker vindt, komt wellicht niet op zo'n site.

Opdracht 4 bladzijde 12

Ann Landers was jarenlang de bezielster van een dagelijkse adviesrubriek in een Amerikaanse krant. Naar aanleiding van een verhaal van ouders met probleemkinderen, lanceerde ze zelf de vraag: 'Mocht u opnieuw kunnen kiezen, zou u dan opnieuw kinderen willen?' Ze kreeg bijna 10 000 antwoorden toegestuurd. Op basis van die bevraging verscheen kort nadien een artikel met als opzienbarende titel: *70 % of parents say 'Kids not worth it'*.

Enige tijd later werd dezelfde vraag door een professioneel onderzoeksbureau aan 1373 personen gesteld. In die steekproef bleek dat 9 % van de ouders spijt had kinderen te hebben.

Leg uit waarom het steekproefresultaat van Ann Landers onbetrouwbaar is en waarom het zoveel hoger is dan het veel betrouwbaarder resultaat van het onderzoeksbureau.

Ten gevolge van vrijwillige respons heeft Ann Landers wellicht vooral brieven ontvangen van ouders die problemen hebben (gehad) en die eindelijk kwijt konden dat ze dit niet opnieuw zouden willen.

Opdracht 5 bladzijde 12

Geef enkele voorbeelden van steekproeven, zelf bedacht of gevonden in de media, waarbij vrijwillige respons of opportunisme de representativiteit van de steekproef betwifelbaar maken.

Opdracht 6 bladzijde 13

Je maakt een doos met honderd niet te onderscheiden briefjes. Op een aantal van die briefjes staat een kruisje, de overige zijn blanco. Indien je met meerdere dozen werkt, moet elke doos hetzelfde aantal gemerkte en blanco briefjes hebben. De briefjes worden goed gemengd. De andere personen in de klas weten niet op hoeveel briefjes een kruisje staat.

Elk om de beurt trekt iedereen uit de klas, zonder te kijken, tien briefjes uit de doos, noteert het aantal dat een kruisje bevat en legt ze terug in de doos, waarna alles opnieuw goed gemengd wordt.

Elke trekking van tien briefjes vormt een EAS uit de populatie van de honderd briefjes. Met elke steekproef komt één steekproefresultaat overeen. We onderzoeken de verdeling van de steekproefresultaten.

- 1 Maak een staafdiagram van de steekproefresultaten.
- 2 Tussen welke waarden variëren de steekproefresultaten? Hoe verklaar je die variatie?
- 3 Maak een schatting voor het vermoedelijke aantal briefjes met een kruisje en vergelijk dit resultaat met het werkelijke aantal in de doos.

Opdracht 7 bladzijde 22

Herneem de 1000 resultaten uit het voorbeeld in het theoriekader.

- 1 Leid het betrouwbaarheidsniveau af dat hoort bij een foutenmarge van 2 %.

$$\frac{55 + 82 + 121 + 151 + 162 + 138 + 94 + 74}{1000} = 0,877 \text{ of } \frac{945 - 33 - 35}{1000} = 0,877,$$

dus 87,7 %

- 2 Je wil meer zekerheid en wil een betrouwbaarheidsniveau van minstens 99,5 %. Welke foutenmarge zou je dan moeten gebruiken?

$$\frac{945 + 16 + 16}{1000} = 0,977 = 97,7 \% \text{ en } \frac{977 + 9 + 10}{1000} = 0,996 = 99,6 \%$$

99,6 % van alle steekproeven ligt binnen een maximale afstand van 0,035 van de werkelijke populatieproportie. De foutenmarge is dus 3,5 %.

Opdracht 8 bladzijde 22

In een krant schrijft een journalist: *'Uit de laatste opiniepeiling blijkt dat 53 % van de stemgerechtigden voor kandidaat A zal stemmen in de volgende presidentsverkiezingen. Deze peiling heeft een foutenmarge van 2 %. Dus zijn we zeker dat kandidaat A zal winnen.'*

- 1 Waar zit de fout in deze verkeerde interpretatie van foutenmarges?

Aangezien geen betrouwbaarheidsniveau wordt vermeld, mag men uitgaan van 95 % betrouwbaarheid. Dit betekent dat een methode werd gevolgd die in 95 % van de gevallen een interval oplevert dat de werkelijke proportie bevat. De redeneerfout bestaat erin te denken dat de kans 100 % is, m.a.w. dat de populatieproportie met zekerheid in het interval $[0,51; 0,55]$ ligt.

- 2 Geef zelf een correct betrouwbaarheidsinterval en leg uit wat de betekenis is van dat interval aan iemand die niets van statistiek weet.

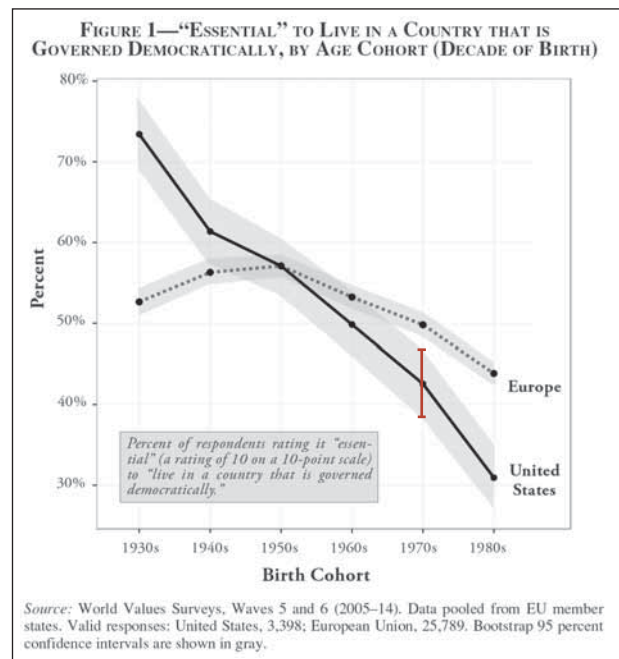
Een juiste uitspraak is: "Het interval $[0,51; 0,55]$ bevat de werkelijke proportie A-stemmers met een betrouwbaarheid van 95 %". Er is bij de gevolgde methode echter een kleine kans dat een interval de populatieproportie niet bevat; dit doet zich voor in hooguit 5 % van de gevallen.

Een andere mogelijke uitspraak is: "Het interval $[0,51; 0,55]$ heeft een kans van 95 % om de werkelijke proportie A-stemmers te bevatten."

Opdracht 9 bladzijde 25

In een ophefmakend artikel over de sterkte van de democratie in Europa en de Verenigde Staten vermelden de auteurs onder andere dat er bij de jongere generaties relatief meer mensen zijn die democratie als niet-essentieel aangeven.

De eerste figuur in hun artikel toont aan dat meer dan 70 % van de Amerikanen die in de jaren 1930 geboren zijn vandaag de dag een democratisch systeem essentieel vindt, terwijl dat bij volwassenen die in de jaren 1980 geboren zijn nog maar rond de 30 % is.



Op de website van de *World Values Survey*, waarnaar de auteurs verwijzen, vonden we voor de peiling van de periode 2010–2014 de volgende cijfers voor het percentage respondenten dat 'absoluut belangrijk' antwoordt op de vraag naar het belang van democratie: 29,2 % van de 488 geïnterviewde personen tot en met 29 jaar; 43,2 % van de 746 personen van 30 tot en met 49; 57,4 % van de 998 personen van 50 of ouder.

1 Geef bij deze drie steekproefproporties een betrouwbaarheidsinterval voor de populatieproportie, bij een betrouwbaarheid van 95 %.

- $\hat{p} = 29,2\% \approx 30\%$ en $n = 488$, zodat $400 < n < 600$.
In de tabel moeten we dus kiezen tussen 4,5 % en 3,7 %. Omdat de betrouwbaarheid van onze uitspraak zeker 95 % zou zijn, gebruiken we de grootste foutenmarge, hier dus 4,5 %.
Het betrouwbaarheidsinterval vinden we als volgt: $0,292 \pm 0,045$ of $[0,247; 0,337]$.
- $\hat{p} = 43,2\%$, zodat $40\% < \hat{p} < 50\%$ en $n = 746$, zodat $600 < n < 800$.
De grootste bijbehorende foutenmarge in de tabel is 4,0 %.
Het betrouwbaarheidsinterval is: $0,432 \pm 0,040$ of $[0,392; 0,472]$.
- $\hat{p} = 57,4\%$, zodat $50\% < \hat{p} < 60\%$ en $n = 998 \approx 1000$.
De grootste bijbehorende foutenmarge in de tabel is 3,1 %.
Het betrouwbaarheidsinterval is: $0,574 \pm 0,031$ of $[0,543; 0,605]$.

2 Leg aan iemand die weinig van statistiek weet uit wat zo'n betrouwbaarheidsinterval betekent.

Het betrouwbaarheidsinterval heeft een kans van 95 % om de werkelijke populatieproportie te bevatten.

Zo is er bijvoorbeeld een kans van 95 % dat de werkelijke proportie personen tot en met 29 jaar dat de democratie absoluut belangrijk vindt, ligt tussen 24,7 % en 33,7 %.

- 3 Op de figuur geven de grijze banden een 95 %-betrouwbaarheidsinterval bij de percentages bij de verschillende geboortejaren. Ter verduidelijking tekenden we in het rood zo'n interval voor het resultaat van de Amerikanen geboren in de jaren 70.

De betrouwbaarheidsintervallen voor Europa zijn blijkbaar een stuk smaller dan die van de Verenigde Staten. Wijst dat op een grotere of op een kleinere steekproef van Europeanen, in vergelijking met die van Amerikanen? Verklaar.

De breedte van het betrouwbaarheidsinterval wordt bepaald door de foutenmarge. De foutenmarge hangt af van de steekproefgrootte: hoe groter de steekproef, hoe kleiner de foutenmarge, dus hoe smaller het betrouwbaarheidsinterval.

De foutenmarge hangt ook af van de populatieproportie. Omdat deze bij de gegeven figuur maximaal 20 % verschillen, verklaart dit het grote verschil in foutenmarge niet.

De steekproef van Europeanen was dus groter.

- 4 Stel dat je foutenmarges wil geven bij een betrouwbaarheid van 99 %. Zullen die groter, kleiner of gelijk zijn aan die bij 95 % betrouwbaarheid?

De foutenmarges f zullen groter zijn opdat het interval $[\hat{p} - f, \hat{p} + f]$ een kans van 99 % heeft om de gezochte populatieproportie te bevatten.

Opdracht 10 bladzijde 30

Bij elk van de vermelde steekproefontwerpen is er minstens één bron van vertekening. Geef zoveel mogelijk bronnen van vertekening.

- 1 Een fruithandelaar heeft een grote kist appels gekocht en test de kwaliteit van de appels door er een tiental die bovenaan liggen te inspecteren.

De bovenste appels vormen een opportunistische steekproef.

- 2 Als eindwerk wil een groep laatstejaarsleerlingen onderzoeken hoe hun leeftijdsgenoten denken over democratie. Via het leerplatform van de school sturen ze een bericht naar alle laatstejaarsleerlingen met de vraag een online vragenlijst in te vullen.

De steekproef zal bestaan uit leerlingen van slechts één school (of scholengroep). Niet alle leerlingen volgen de berichten op het leerlingenplatform op en niet iedereen is bereid een online vragenlijst in te vullen. Er is geen garantie dat de vragen ernstig beantwoord zullen worden.

- 3 Om de tevredenheid van zijn telefoonabonnees te achterhalen, belt een telecomoperator 1000 lukraak gekozen telefoonnummers uit zijn eigen klantenbestand op tussen 16 uur en 21 uur en dat op een maandag en dinsdag.

Er is geen garantie dat iedereen bereikbaar is op die momenten. Niet iedereen is bereid zo'n tevredenheidsenquête uit te voeren.

Opdracht 11 bladzijde 30

In de onderstaande uitspraken wordt een kenmerk gegeven van een bepaalde groep mensen of dingen, die *cursief* zijn weergegeven. Geef telkens aan of, volgens jou, *alle* mensen of dingen uit die groep zijn onderzocht of slechts een steekproef.

- 1 Slechts 19 % van de *Vlaamse kinderen tot 6 jaar* voldoet aan de drie voorwaarden om tandbederf maximaal tegen te gaan: beginnen met poetsen voor de eerste verjaardag, minstens twee keer per dag poetsen en hulp van papa of mama bij het poetsen.

Steekproef.

- 2 Het studierendement is de verhouding van het aantal verworven studiepunten (waarvoor geslaagd) t.o.v. het aantal opgenomen studiepunten (waarvoor ingeschreven). Het gemiddelde studierendement van *leerlingen uit een richting Humane Wetenschappen* voor de bachelor Toegepaste Psychologie Algemene Opleiding bedraagt 75,75 %.

Omwille van de formulering kan het enkel om een steekproef gaan: er wordt niet vermeld in welk jaar die leerlingen afstudeerden.

- 3 Uit gegevens van de KU Leuven blijkt dat 26,6 % van de *generatiestudenten in de academische bachelor psychologie die in het academiejaar 2013-2014 aan alle examens deelnamen* een studierendement (zie vorige vraag) van 100 % haalde.

Volledige populatie: de KULEuven beschikt over het studierendement van al hun studenten van deze studierichting in dat academiejaar, evenals van hun geboortjaar.

- 4 Coca-Cola bevat 9,7 mg cafeïne per 100 ml.

Steekproef: slechts enkele stalen worden gecontroleerd.

Opdracht 12 bladzijde 31

Hoe de 'wetenschappelijke methode' om steekproeven te nemen haar ingang vond

Voor de Amerikaanse presidentsverkiezingen van 1936 hadden het enquêtebureau *American Institute of Public Opinion* en het magazine *Literary Digest* op basis van een steekproef elk een voorspelling uitgebracht. De voorspellingen spraken elkaar echter tegen.

Volgens *Literary Digest*, dat sinds 1916 steevast de winnaar correct had voorspeld en daardoor een stevige reputatie genoot, zou de republikein Alf Landon winnen, met ongeveer 57 % van de stemmen. Het pas opgerichte *American Institute of Public Opinion* van George Gallup voorspelde daarentegen een overwinning voor de democraat Franklin D. Roosevelt, die volgens het enquêtebureau zo'n 54 % van de stemmen zou wegkapen.



Literary Digest had duizenden mensen in dienst, die niet minder dan 10 miljoen enquêteformulieren het land rondstuurden. De voornaamste bronnen voor adressen waren hun eigen abonneelijst, een databank van auto-eigenaars (via hun nummerplaatregistratie) en telefoonboeken. Ongeveer 2,4 miljoen aangeschrevenen reageerden, waardoor deze peiling een van de grootste was uit die tijd. Het *Institute* zond slechts een paar honderd interviewers uit, die samen zo'n 3000 Amerikanen persoonlijk interviewden. Elke interviewer moest zich beperken tot een vastgelegd aantal van elke type kiezer: zoveel vrouwen, zoveel blanken, zoveel uit die en die inkomenscategorie, zoveel tussen 30 en 40 jaar oud enzovoort.

Uiteindelijk won Roosevelt glansrijk de verkiezingen met 62,5 % van de stemmen.

- 1 Geef bij beide onderzoeken de populatie en de steekproef.

Populatie: stemgerechtigde Amerikanen (ter informatie: het stemrecht voor blanke vrouwen werd al in 1920 ingevoerd).

Steekproef *Literary Digest*: 2,4 miljoen vrijwillige deelnemers, uit een lijst van 10 miljoen gecontacteerde abonnees, auto-eigenaars, telefoonbezitters en andere.

Steekproef *Institute*: ongeveer 3000 Amerikanen, representatief samengesteld op het gebied van geslacht, huidskleur, inkomsten, leeftijd ... (Volgens sommige internet-pagina's gebruikte George Gallup een steekproef die groter was; de cijfers lopen uiteen, maar feit is wel dat zijn steekproef veel kleiner was dan die van *Literary Digest*.)

- 2 Wat zijn volgens jou enkele zwakke punten in de aanpak van *Literary Digest*?

Ze 'beperkten' zich tot hun eigen abonnees (dat duidelijk uit geschoolde mensen be-

stond) en mensen die zich een auto of telefoon konden veroorloven. Bovendien be-
dienden ze zich van een steekproef op basis van vrijwillige respons: het is zeer goed
mogelijk dat vooral mensen die verandering wilden (Roosevelt was op dat ogenblik al
president) hun alternatieve voorkeur bekend maakten.

- 3 Welke partij telde onder haar kiezerspubliek de meeste gegoede burgers?

Toen telden de Republikeinen veel gegoede burgers in hun kiespubliek.

- 4 Geef enkele mogelijke verklaringen voor het feit dat *Literary Digest* de bal compleet missloeg.

Ze gebruikten een vertekende steekproef: Republikeinen hadden een grotere kans om
aangeschreven te worden en mogelijk kwamen vooral diegenen die verandering wilden
in de uiteindelijke steekproef terecht.

Dit historische voorbeeld geeft goed aan dat een slechte steekproef niet gecorrigeerd kan
worden door grote aantallen (in dit geval ongeveer 2,4 miljoen respondenten).

Opdracht 13 bladzijde 32

In 1987 verscheen het ophefmakende boek *'Women and Love: A Cultural Revolution in Progress'*
van Shere Hite. Uit onderzoek van de auteur, uitsluitend bij vrouwen, bleek dat van de
respondenten:

- 84 % zich emotioneel niet goed voelde in hun relatie,
- 95 % psychische of lichamelijke mishandeling door hun partner had ondergaan,
- slechts 13 % na twee jaar huwelijk nog aangaf van hun man te houden en
- 70 % van de vrouwen die langer dan 5 jaar gehuwd waren, buitenechtelijke relaties had.

Hite onderbouwde de representativiteit van haar onderzoek voor de hele Amerikaanse bevolking
met de onderstaande tabellen:

Jaarlijks inkomen (x 1000 dollar)		
klasse	steekproef (%)	V.S. (%)
[0, 2[19,0	18,3
[2, 6[22,5	25,4
[6, 10[17,0	17,1
[10, 15[13,0	15,0
[15, 20[10,0	9,8
[20, 25[8,0	6,4
≥ 25	8,5	8,2

Woonplaats		
	steekproef (%)	V.S. (%)
Grootsteden	60	62
Platteland	27	26
Kleine steden	13	12

Ras / afkomst		
	steekproef (%)	V.S. (%)
Blank	82,5	83,0
Zwart	13,0	12,0
Aziatisch	1,8	2,0
Spaans	1,8	1,5
Indiaan	0,9	1,0
Midden-Oosten	0,3	0,5

Regio		
	steekproef (%)	V.S. (%)
Oosten	21	22
Midden	27	26
Zuiden	31	33
Westen	21	19

Dit onderzoek is een mooi en waar gebeurd voorbeeld van hoe het niet moet. Nochtans
zijn een aantal aspecten van de steekproef zeer bekoorlijk: het is een vrij grote steekproef
(n = 4500) en ogenschijnlijk geeft hij een waarheidsgetrouwe doorsnede van de Ameri-
kaanse vrouwelijke volwassen bevolking weer.

- 1 Wat denk je van de betrouwbaarheid van de steekproef (n = 4500), in het licht van de
bovenstaande informatie?

Al zullen een aantal leerlingen wellicht onraad ruiken, velen zullen geneigd zijn dit als een betrouwbare, want 'representatieve' steekproef te beschouwen. Het extreme van de bevindingen wekt uiteraard ook argwaan.

- 2 De vragenlijst bevatte meer dan 100 vragen. De respondenten moesten die zelf thuis invullen en terugsturen. Shere Hite had ook interviewers kunnen inschakelen om de vragenlijst via een persoonlijk gesprek te laten invullen. Zou dat betrouwbaarder geweest zijn bij een dergelijk onderzoek naar relaties en seksualiteit?

Bij onderzoek naar gevoelige en persoonlijke onderwerpen, zoals relaties en seksualiteit, wordt in de praktijk weinig met interviewers gewerkt, omdat daardoor een grote vertekening in de antwoorden kan ontstaan: mensen zijn dan vaak geneigd het maatschappelijk aanvaardbare antwoord te geven of minder fraaie aspecten van zichzelf te verzwijgen. Dit euvel kan verholpen worden via enquêtes per computer: een enquêteur levert de respondent een computer met een softwaregestuurde vragenlijst, onder de garantie dat de gegevens anoniem worden verzameld en verwerkt. Een papieren vragenlijst kan ook op een gelijkaardige manier voor anonimiteit zorgen.

Besluit: een enquête via een vragenlijst is in dit onderzoek betrouwbaarder dan een bevraging door een interviewer.

- 3 In 1987 verscheen een opiniepeiling, uitgevoerd door de firma Harris, waaruit bleek dat 89 % van de respondenten wel degelijk voldoening putte uit hun relatie; dit cijfer was ongeveer gelijk voor mannen en vrouwen. Volgens Hite voelden 84 % van de vrouwen zich niet goed in hun relatie.

Welk cijfer ligt het dichtst bij de werkelijkheid, denk je?

Hopelijk worden de resultaten van de firma Harris unaniem als waarschijnlijker aangeduid ...

- 4 Hite had meer dan 100 000 vragenlijsten rondgestuurd via vrouwenorganisaties en andere kanalen. Zo'n 4500 vrouwen vulden de vragenlijst in. Ga na of dit gevolgen heeft voor de betrouwbaarheid van de steekproef: welke groep vrouwen zal eerder wel of eerder niet reageren? Wie heeft geen kans om te reageren?

Je kunt verwachten dat vrouwen met een zorgeloze relatie minder gemotiveerd zullen zijn om de lange vragenlijst in te vullen en per post terug te sturen, dan diegene met heel nare relationele ervaringen, die via deze bevraging een mogelijkheid zien om 'hun verhaal' of zelfs hun 'noodkreet' de wereld in te sturen. Indien dit inderdaad zo is, dan toont de ogenschijnlijke representativiteit van de steekproef aan dat dergelijke relatieproblemen in alle lagen van de bevolking, los van ras of herkomst en op alle plaatsen in de V.S. voorkomen.

Door de enquêtes te versturen via vrouwenorganisaties en andere kanalen specifiek voor vrouwen, hebben vrouwen die daar niet bij aangesloten zijn minder kans om zo'n enquête in handen te krijgen. Een belangrijke vraag is dan of vrouwenorganisaties bijvoorbeeld een representatief staal van de vrouwelijke populatie bereiken.

Opdracht 14 bladzijde 33

Een betoging zegt iets over de mening van de bevolking. De manifestanten zijn te beschouwen als een steekproef, waaruit je besluiten over de hele populatie kunt trekken. Politici worden geacht om op dergelijke signalen van de bevolking passend te reageren.

Aan welke bekende vorm van vertekening lijdt een betoging echter?

Hier is duidelijk sprake van vrijwillige respons.

Opdracht 15 bladzijde 33**Stemplicht of stemrecht?**

Sommigen pleiten voor het afschaffen van de kiesplicht, ten voordele van kiesrecht, zoals in de meeste van de ons omringende landen. Eén van de argumenten daarbij is soms dat extreme partijen dan minder kiezers zullen hebben: mensen met een negatief beeld of zelfs een aversie van politiek en politici, zo gaat de redenering, zullen dan wellicht thuisblijven i.p.v. uit protest op extreme partijen te stemmen.

Bij kiesrecht kun je een verkiezing beschouwen als een steekproefname, aangezien slechts een deel van de populatie zijn mening kenbaar maakt.

Vind jij het hierboven geschetst argument vóór de afschaffing van de kiesplicht overtuigend? Waarom (niet)?

Ten gevolge van de vrijwillige respons zou het evenzeer kunnen dat vooral mensen die de regerende politici beu zijn en die radicale veranderingen wensen, gaan stemmen, wat dus voor een toename van het aantal extreme stemmen zou kunnen zorgen.

Opdracht 16 bladzijde 33**Andere bronnen van vertekening**

Ondanks een zorgvuldig samengestelde steekproef, kan in een peiling nog altijd vertekening aanwezig zijn. Veel belang moet gehecht worden aan de afwezigheid van enige beïnvloeding van de geënquêteerde. De interviewer mag dus geen appreciatie of afkeuring laten blijken bij zijn/haar antwoorden; ook moeten de vragen zorgvuldig geformuleerd worden.

Vergelijk de verschillende vragen en geef aan welke volgens jou het meest betrouwbare antwoord zullen opleveren.

- 1 'Kijk je naar tekenfilms?', 'Kijk je nog altijd naar tekenfilms?' of 'Studies tonen aan dat mensen die naar tekenfilms kijken slechter presteren op school. Kijk jij naar tekenfilms?'

Hier is vrij duidelijk dat de eerste vraag het meest waarde-neutraal is. De toevoeging 'nog altijd' in de tweede formulering geeft aan dat het niet echt hoort om 'nog' naar tekenfilms te kijken op een bepaalde leeftijd. Ook de derde versie plaatst naar tekenfilms kijken in een kwalijk daglicht.

- 2 'Moet het begrotingsoverschot gebruikt worden voor een belastingsverlaging of om nieuwe overheidsinitiatieven op te starten?' of 'Moet het begrotingsoverschot gebruikt worden voor een belastingsverlaging of geïnvesteerd worden in nieuwe initiatieven voor onderwijs, milieu, gezondheidszorg, bestrijding van criminaliteit en dergelijke?'

Hier is moeilijk uit te maken welke vraag het meest betrouwbare antwoord zal opleveren. Maar afhankelijk van welk antwoord je wil, kun je via een van beide vragen zowel steun voor of weerstand tegen belastingverlagingen 'statistisch aantonen'. In de eerste vraag is 'belastingvermindering' het meest concrete voorstel, in de tweede komen de alternatieven beter uit de verf.

3 'Merkt u ook dat de onveiligheid sterk is toegenomen?' of 'Voelt u zich veilig in uw stad?'

In de eerste vraag suggereert de formulering 'merkt u ook' dat vele anderen dit al opmerkten, zodat de geënquêteerde aangezet wordt ook 'ja, ik ook' te antwoorden. De tweede vraag is minder suggestief.

Opdracht 17 bladzijde 34

Onderzoek naar het gemiddelde aantal kinderen in een gezin

- 1 Hoeveel kinderen telt je gezin, jezelf inbegrepen? Broers en zussen die niet meer thuis wonen, moet je ook meetellen. Verzamel die cijfers voor iedereen van je klas en bepaal het gemiddelde.
- 2 Om na te gaan of het gemiddelde voor jouw klas een betrouwbare schatting is van het gemiddelde in Vlaanderen, is de volgende vraag van belang. Stel dat in je school aan een EAS van leerlingen wordt gevraagd hoeveel kinderen er in hun gezin zijn, henzelf inbegrepen. Zullen alle gezinnen dan dezelfde kans hebben om geselecteerd te worden? Met andere woorden, geeft een EAS van leerlingen aanleiding tot een EAS van gezinnen (= de onderzochte eenheid)?

Gezinnen zonder kinderen kunnen onmogelijk in deze steekproef terechtkomen. Het gemiddeld aantal kinderen per gezin zal dus lager liggen.

Opdracht 18 bladzijde 34

Elk jaar zijn er heel wat katten die uit ramen of van balkons van flatgebouwen vallen. Dierenartsen spreken van HRS of *feline high-rise syndrome*. Het is daarbij opmerkelijk dat katten hun val vaak overleven. Sterker nog: uit een onderzoek in samenwerking met een dierenkliniek in New York bleek dat katten die van hoger vallen een grotere kans hebben om te overleven!

Het onderzoek liep van 4 juni 1984 tot 4 november van dat jaar. In die periode werden 132 gewonde katten binnengebracht. Daarvan kregen er 17 een spuitje, omdat hun baasjes de behandeling niet konden betalen en 8 van de 115 overblijvende katten (of 7,0 %) stierven ten gevolge van hun val. Van de 22 katten die meer dan zeven verdiepingen naar beneden tuimelden, stierf er maar 1 (4,8 %). Bij de 13 die negen verdiepingen vielen, was de ergste verwonding een botbreuk (0 % sterfgevallen dus) en een kat die tweeëndertig verdiepingen viel, had een lichte longperforatie en een gebroken tand (0 % sterfgevallen).

Hoewel er een aantal mogelijke verklaringen zijn voor deze merkwaardige overlevingskansen, moet eerst een andere vraag beantwoord worden: is deze steekproef van 132 katten betrouwbaar voor de populatie 'katten die uit flatgebouwen vallen'? Met andere woorden, hadden alle katten die (in de buurt van die dierenkliniek) uit een flatgebouw vielen evenveel kans om in de steekproef van dat onderzoek te belanden?

De achtergrond voor deze opdracht staat op www.cut-the-knot.org/do_you_know/misuse.shtml waar verwezen wordt naar het boek van Marilyn vos Savant 'The power of logical thinking'. Daarin geeft de auteur toe dat ze oorspronkelijk zelf niet had stilgestaan bij het feit dat enkel de overlevende katten in de statistieken van die kliniek konden terechtkomen. Katten die onmiddellijk sterven worden niet naar een dierenarts gebracht. De steekproef is dus vertekend.

Opdracht 19 bladzijde 35**Project 'Bronnen van vertekening'**

Het opportunistisch samenstellen van een steekproef, evenals vrijwillige respons, kan voor vertekening van die steekproef zorgen. Maar er zijn heel wat andere bronnen van vertekening.

Zoek in handboeken statistiek of op het internet een aantal andere bronnen en groepeer ze in twee grote groepen:

- vertekeningen die ontstaan bij het *samenstellen* van de steekproef;
- vertekeningen die ontstaan bij het *afnemen van de enquête*.

Bij het zoeken kan het nuttig zijn enkele Engelse vertalingen te kennen. Steekproef: *sample*; vertekening: *bias*; opiniepeiling: *survey of poll*; opportunistische steekproef: *convenience sample*; steekproef op basis van vrijwillige respons: *voluntary response sample*; enkelvoudig aselechte steekproef (EAS): *simple random sample (SRS)*.

Opdracht 20 bladzijde 35**Project 'Praktische alternatieven voor een EAS'**

Theoretisch gezien is een enkelvoudig aselechte steekproef interessant. In de praktijk worden echter andere soorten steekproeven gebruikt, die gemakkelijker en goedkoper samen te stellen zijn. Sommige leveren ook betrouwbaarder resultaten op dan een even grote EAS.

Verzamel informatie i.v.m. *gestratificeerde steekproeven* en *geblokte steekproeven* en vergelijk die (voor- en nadelen) met enkelvoudig aselechte steekproeven.

Opdracht 21 bladzijde 35**Project 'Peilen naar zeer persoonlijke kwesties'**

Bij enquêtes rond gevoelige onderwerpen, zoals drugsgebruik, emoties, seksualiteit of inkomens, is het niet gemakkelijk om betrouwbare antwoorden te krijgen. Mensen zijn namelijk geremd om eerlijk te antwoorden op vragen die ze als zeer persoonlijk ervaren.

Daarom gebruiken onderzoekers in die gevallen soms een techniek die in het Engels *randomized response* genoemd wordt: het antwoord (*response*) wordt door het toeval (*random*) bepaald. Een deel van de geënquêteerden antwoordt daarbij op een onschuldige vraag, die niets met het onderzoek te maken heeft, en de anderen antwoorden op de vraag die wel van belang is, maar zó dat de onderzoeker niet kan weten op welke van beide vragen de respondent antwoordt.

Zoek in externe bronnen minstens één methode van *randomized response*. Illustreer ze aan de hand van een concrete situatie, fictief of waargebeurd.

Opdracht 22 bladzijde 36**Project 'Een statistisch rapport evalueren'**

Elk onderzoek wil je overtuigen, maar sommige onderzoeken hebben daar meer argumenten voor dan andere. Het is nuttig om snel het kaf van het koren te kunnen scheiden.

Zoek met jouw projectgroep een drietal rapporten van statistisch onderzoek, op het internet bijvoorbeeld, en probeer aan de hand daarvan zoveel mogelijk van de onderstaande vragen te beantwoorden. Zorg dat er minstens één onderzoek bij zit waarin uitdrukkelijk wordt gerapporteerd over de gevolgde methode bij het samenstellen van de steekproef. Dat gebeurt vaak helemaal vooraan of helemaal achteraan in het rapport, onder een hoofding 'methodologie', 'methode' ...

- **Wie voerde het onderzoek uit?**
Een bekend peilingbureau heeft een reputatie hoog te houden; een kleine organisatie die ijvert voor een of andere maatschappelijke, economische, ecologische ... verandering zal het minder nauw nemen met de vereisten van goed statistisch onderzoek.
- **Wat was de beoogde populatie?**
Wilden de onderzoekers de opinie van jongeren? Van alle Vlamingen (en dus ook kinderen)? Van Europeanen van 18 of ouder?
- **Hoe werd de steekproef samengesteld?**
Een degelijk onderzoek vermeldt altijd nauwgezet hoe een steekproef werd samengesteld. Dat is een essentieel onderdeel van elk betrouwbaar statistisch rapport. Die informatie wordt echter zelden vermeld in krantenartikels of op websites.
- **Hoe groot was de steekproef?**
Grotere steekproeven zijn betrouwbaarder, indien ze goed zijn samengesteld. Het is dus belangrijk de steekproefgrootte precies te kennen.
- **Wat was de deelnemingsgraad?**
Het is zelden zo dat alle personen uit de opgestelde steekproef willen/kunnen deelnemen: we spreken dan van non-respons. Indien te weinig mensen meewerkten of, met andere woorden, indien de non-respons te groot is, zijn resultaten minder betrouwbaar.
- **Hoe gebeurde het interview?**
Betrof het een telefonisch interview, een schriftelijke enquête, een mondelinge bevraging? Voor waardegeladen onderwerpen is een mondeling interview minder opportuun, omdat mensen niet altijd eerlijk durven te antwoorden. In dergelijke situaties zijn anonieme schriftelijke of computergestuurde enquêtes een oplossing. Maar in andere gevallen is een persoonlijk interview dan misschien weer betrouwbaarder.
- **Wanneer vond de peiling plaats?**
Recente gebeurtenissen kunnen de mening van mensen beïnvloeden: een dramatische moord, een ongeval, een beurskrach ...
- **Wat waren de exacte vragen?**
De verwoording van de vraag kan het antwoord sterk in een bepaalde richting sturen.

Opdracht 23 bladzijde 37

Uit een onderzoek bij 1000 Vlamingen, uitgevoerd begin september 2016, bleek dat 91,9 % van de Vlamingen vindt dat de dotaties van prinses Astrid en prins Laurent beperkt of afgeschaft moeten worden als ze hun taken niet naar behoren vervullen. Het onderzoeksbureau iVOX, dat het onderzoek uitvoerde, vermeldt bij de peiling dat de foutenmarge 3,02 % bedraagt, zonder een betrouwbaarheids-niveau te geven.

- 1 Leg aan iemand die weinig van statistiek weet uit wat die foutenmarge betekent.

Aangezien geen betrouwbaarheidsniveau wordt vermeld, mogen we uitgaan van 95 % betrouwbaarheid.

Een foutenmarge van 3,02 % betekent dat de kans 95 % is dat de steekproefproportie niet meer dan 3,02 % afwijkt van de gezochte populatieproportie.

Hier is er dus een kans van 95 % dat de 91,9 % uit het onderzoek niet meer dan 3,02 % afwijkt van de werkelijke proportie Vlamingen die vindt dat de dotaties van prinses Astrid en prins Laurent beperkt of afgeschaft moeten worden als ze hun taken niet naar behoren vervullen.

- 2 De foutenmarge die iVox opgeeft, wijkt af van wat je uit de tabel met foutenmarges op p. 24 kunt afleiden. Wat zou dit verschil kunnen verklaren?

Ze werken hoogst waarschijnlijk niet met een enkelvoudig aselechte steekproef, waardoor ze ook met andere foutenmarges moeten werken.

Opdracht 24 bladzijde 37

Transgenders zijn mensen bij wie het geboortegeslacht niet overeenstemt met hoe ze zich voelen. In 2015 werd een aantal Vlamingen bevraagd over hun houding tegenover transgenders. Er werd een steekproef gehouden bij 1500 meerderjarige Nederlandstalige inwoners van het Vlaams en Brussels Hoofdstedelijke Gewest. Zo'n 66 % van de respondenten is het eens met de stelling 'Als iemand goed heeft nagedacht over het veranderen van zijn of haar geslacht, dan is een operatie een goed idee.'

Welke benaderende foutenmarge kun je gebruiken bij een betrouwbaarheid van 95 %?

$\hat{p} = 66\%$, zodat $60\% < \hat{p} < 70\%$ en $n = 1500$.

De grootste bijbehorende foutenmarge in de tabel is 2,5 %.

Opdracht 25 bladzijde 37

Bij een EAS van 500 kotstudenten bleek dat er 44 zich met een elektrische fiets verplaatsen van en naar hun universiteit of hogeschool.

Bepaal een 95 %-betrouwbaarheidsinterval.

$\hat{p} = \frac{44}{500} = 0,088 = 8,8\% \approx 10\%$ en $n = 500$, zodat $400 < n < 600$.

De grootste bijbehorende foutenmarge in de tabel is 2,9 %.

Het betrouwbaarheidsinterval is: $0,088 \pm 0,029$ of $[0,059; 0,117]$.

Opdracht 26 bladzijde 37

In de aanloop van de verkiezingen doet een firma wekelijks een opiniepeiling via een EAS van 2000 personen. De laatste dagen vóór de verkiezingen verhoogt ze echter haar steekproef naar 4000.

- 1 Zorgt een grotere steekproef voor minder vertekening? Verantwoord.

Een grotere steekproef zorgt niet voor minder vertekening. Immers, indien men werkt met een EAS, zal de vertekening niet beïnvloed worden door de steekproefgrootte.

Indien men heel veel steekproeven zou nemen van 2000 personen, zou men als gemiddelde van alle gevonden steekproefproporties een waarde vinden die ongeveer gelijk is aan de werkelijke populatieproportie. Dit zou ook het geval zijn, indien men heel veel steekproeven zou nemen van 4000 personen.

- 2 Zorgt een grotere steekproef voor een grotere nauwkeurigheid (kleinere variabiliteit)? Verklaar dit met behulp van de formule voor de standaardafwijking.

Uit simulaties kan vastgesteld worden dat grotere steekproeven een kleinere variabiliteit vertonen. Wanneer men heel veel steekproeven van 4000 personen neemt, zullen de gevonden steekproefproporties \hat{p} minder van p afwijken, dan wanneer men steekproeven van 2000 personen neemt. Zie ook de afbeeldingen op p. 23 en de tabel op p. 24.

Opdracht 27 bladzijde 38

Wanneer je een muntstuk n keer opgooit en dan de relatieve frequentie van het aantal keer kop bepaalt, dan kun je dit experiment interpreteren als het nemen van een steekproef van grootte n uit de oneindige populatie van worpen met dat muntstuk. De kans op kop is de populatieproportie p . De relatieve frequentie van kop is de steekproefproportie \hat{p} .

Je wil de eerlijkheid van een muntstuk nagaan door het heel vaak op te gooien. Hoe vaak moet je dat minstens doen, als je wil dat de gevonden relatieve frequentie van het aantal keer kop hoogstens 2,5 % afwijkt van de werkelijke kans? (Geef het aantal dat in 95 % van de gevallen aan deze voorwaarde zal voldoen.)

Indien je 1500 keer simuleert, is de foutenmarge voor alle mogelijke kansen p (dus ook voor een populatieproportie van rond de 50 %) kleiner dan 2,5 % (met een betrouwbaarheid van 95 %).

Opdracht 28 bladzijde 38

Onderzoeken wijzen uit dat ongeveer 15 % van de Belgische mannen ouder dan 30 jaar last heeft van een verhoogde bloeddruk. In een bepaald gebied blijkt water besmet te zijn door een chemisch bedrijf in die omgeving. Onderzoekers denken dat mannen boven de 30 hierdoor gevoeliger zijn voor een hoge bloeddruk. In een EAS van 300 mannen blijken er 72 een hogere bloeddruk te hebben.

Bereken met deze gegevens een 95 %-betrouwbaarheidsinterval. Wat denk je zelf over de geloofwaardigheid van het verband tussen het besmette drinkwater en het hebben van een te hoge bloeddruk?

$$\hat{p} = \frac{72}{300} = 0,24 = 24 \%, \text{ zodat } 20 \% < \hat{p} < 30 \%$$

$$n = 300, \text{ zodat } 200 < n < 400.$$

De grootste bijbehorende foutenmarge in de tabel is 6,4 %.

Het betrouwbaarheidsinterval is: $0,240 \pm 0,064$ of $[0,176; 0,304]$.

Met een betrouwbaarheid van 95 % ligt het percentage mannen in dat gebied dat last heeft van een hoge bloeddruk tussen 17,6 % en 30,4 %. Dit is beduidend meer dan de gegeven 15 %. Maar dit statistisch resultaat laat ons niet toe de oorzaak te achterhalen. Het kan wel een aanzet zijn tot verder onderzoek.

Opdracht 29 bladzijde 39**Simulaties met een grafisch rekentoestel**

In een bepaalde populatie heeft een kenmerk een proportie $p = 0,23$.

Je kunt het nemen van een steekproef van 50 personen simuleren door 50 lukrake gehele getallen van 1 tot en met 100 te genereren. De nummers 1 tot en met 23 associëren we met mensen 'met het kenmerk' en de nummers vanaf 24 met die 'zonder het kenmerk'.

Het commando **randInt(1,100,50)** genereert 50 lukrake gehele getallen van 1 tot en met 100. Een mogelijk resultaat is {18 2 54 8 84 85 ...}.

Breiden we **randInt(1,100,50)** uit met de logische test ≤ 23 , dan wordt van elk van de 50 getallen getest of het kleiner is dan of gelijk aan 23. Is dat zo, dan wordt op die plaats een 1 geplaatst ('true'), anders komt er een 0 ('false'). De kans op een 1 is telkens 0,23. Met de getallen uit de eerste stap zou **randInt(1,100,50) ≤ 23** als resultaat geven: {1 1 0 1 0 0 ...}.

Het commando **sum()** zorgt er tot slot voor dat al die enen en nullen opgeteld worden. Je krijgt daarmee de absolute frequentie van het aantal enen, met andere woorden het aantal personen in de steekproef die het onderzochte kenmerk hebben. Je vindt de steekproefproportie dus als **sum(randInt(1,100,50)≤23)/50**.

Je kunt dergelijke simulatie van één steekproef 200 keer herhalen door gebruik te maken van het commando: **seq(sum(randInt(1,100,50)≤23)/50,X,1,200)**. De variabele X neemt achtereenvolgens de waarden 1 tot en met 200 aan en voor elk van die waarden wordt de simulatieopdracht (in het blauw) opnieuw uitgevoerd. Een mogelijk resultaat is **{.3 .24 .34 .32 .2 .24 ...}**.

- 1 Voer de bovenstaande simulatie uit en stel de steekproefresultaten voor in een histogram.
- 2 Simuleer nu op een analoge manier heel wat steekproeven van 100 worpen met een eerlijk muntstuk en geef een schatting van de kans dat $\hat{p} > 0,55$.

Opdracht 30 bladzijde 40

Hieronder worden enkele peilingen beschreven. Geef telkens de populatie, de steekproef en de eventuele bronnen van vertekening.

- 1 In de finale van een televisiewedstrijd voor beginnend Vlaams zangtalent mag elke Vlaming van 12 jaar of meer de winnaar helpen verkiezen. Er kan enkel gestemd worden per (betalende) sms, één sms per telefoonnummer, tussen 21 uur en 21.30 uur. In totaal brengen 323 641 personen hun stem uit.

Populatie: elke Vlaming van 12 jaar of meer. **Steekproef:** de 323 641 personen die een sms uitbrengen. **Vertekeningen:** niet iedereen heeft een gsm, niet iedereen wil hier geld aan uitgeven, wellicht zijn massaal vrienden en familie opgetrommeld ...

- 2 Een gemeentebestuur wil het gevoel van veiligheid bij de bevolking verhogen en wil in eerste instantie via een bevraging vernemen welke de belangrijkste knelpunten zijn. Er wordt een enkelvoudig aselechte steekproef van 500 adressen uit de adressenlijst van die stad gekozen. De buurtpolitie krijgt als taak op die adressen te gaan aanbellen en de eerste persoon die opendoet en die 18 is of ouder de vragenlijst van de enquête voor te leggen.

Populatie: elke inwoner van de gemeente. **Steekproef:** de personen die als eerste openen en die minstens 18 jaar oud zijn, op de 500 geselecteerde adressen. **Vertekeningen:** afhankelijk van het uur waarop de buurtagenten aanbellen, zullen mensen die uit werken zijn een kleinere kans hebben om in de steekproef terecht te komen. Ook studerende kinderen zijn tijdens de week mogelijk afwezig.

Opdracht 31 bladzijde 40

TiJdens de Tweede Wereldoorlog werden de Amerikaanse bommenwerpers hevig beschoten door het Duitse leger. Sommige vliegtuigen stortten neer, andere slaagden erin terug te keren ondanks de beschietingen. Het leger wilde de overlevingskansen verhogen, door extra bepantsering aan de toestellen toe te voegen. Aangezien dat het gewicht van het vliegtuig snel doet toenemen, wilden ze er zo weinig mogelijk van gebruiken, maar dan wel op de juiste plaatsen. Om die te weten te komen, werden de kogelinslagen van de beschoten vliegtuigen grondig in kaart gebracht. De minste gaten werden ter hoogte van de cockpit en de motoren gevonden, en de meeste in de romp, de vleugels en de brandstoftanks.



Het was voor de militaire top dan ook logisch om op die meest geraakte plaatsen de extra bepantsering toe te voegen. Maar Abraham Wald, een van de wiskundigen die door de Amerikaanse overheid waren ingeschakeld om statistische militaire problemen op te lossen, dacht daar anders

over. Hij beweerde stellig, en kon dat ook bewijzen, dat de enige plekken die verstevigd moesten worden de motoren en de cockpit waren, met andere woorden de plaatsen met de minste kogelinslagen.

Zelfs zonder statistische berekeningen kun je al enigszins begrijpen waarom Wald waarschijnlijk gelijk heeft, door je de vraag te stellen of de steekproef van de onderzochte vliegtuigen representatief was voor de totale populatie van beschoten vliegtuigen. Hadden alle beschoten vliegtuigen een even grote kans om in de steekproef terecht te komen?

Blijkbaar verhinderden de kogelinslagen in romp, vleugels en brandstoftanks niet dat de vliegtuigen terug konden vliegen naar de luchtmachtbasis, terwijl dat nauwelijks mogelijk was wanneer de motoren geraakt waren, of de cockpit.

De onderzochte vliegtuigen vormden dus geen representatieve steekproef van 'beschoten vliegtuigen'.

Opdracht 32 bladzijde 41

Uit een bevraging, uitgevoerd door een peilingfirma in 2015, bleek dat 45 % van de 1005 bevroagde Amerikanen geloofde dat buitenaardse wezens ooit de Aarde bezochten. Veronderstel dat met een EAS is gewerkt.

Gebruik dit steekproefresultaat om een statistisch correcte uitspraak te doen over de populatieproportie Amerikanen die hierin geloven.

$\hat{p} = 45\%$, zodat $40\% < \hat{p} < 50\%$ en $n = 1005 \approx 1000$.

De grootste bijbehorende foutenmarge in de tabel is 3,1 %.

Het betrouwbaarheidsinterval is: $0,450 \pm 0,031$ of $[0,419; 0,481]$.

Met een betrouwbaarheid van 95 % ligt het percentage Amerikanen dat gelooft dat buitenaardse wezens ooit de Aarde bezochten, tussen 41,9 % en 48,1 %.

Opdracht 33 bladzijde 41

Voor een opiniepeiling over de Belgische bevolking worden in de tien provincies enkelvoudig aselechte steekproeven gekozen. Elke provincie heeft echter een verschillend aantal inwoners. Zal dit gegeven de nauwkeurigheid van de schatting beïnvloeden indien

1 in elke provincie 1000 personen worden ondervraagd?

Neen: variabiliteit wordt beïnvloed door de populatieproportie p en de steekproefgrootte n , maar niet door de populatiegrootte.

2 in elke provincie 0,1 % van de bevolking wordt ondervraagd?

Ja, in dit geval zal de nauwkeurigheid wél anders zijn van provincie tot provincie, omdat de steekproefgrootte nu wel varieert. Al kunnen bijbehorende wijzigingen in p voor een lichte compensatie zorgen.

Opdracht 34 bladzijde 41

Twee onderzoekers, A en B, onderzochten onafhankelijk van elkaar eenzelfde kenmerk aan de hand van een EAS van 1000 eenheden. A beweert dat de populatieproportie 43,2 % bedraagt met een foutenmarge van 3,1 % bij een betrouwbaarheid van 95 %.

- 1 B vond een steekproefproportie van 48,5 %. Mogen we daaruit besluiten dat hij bij die 5 % behoort die verder dan 3,1 % van de werkelijke populatieproportie p ligt?

Neen. Enerzijds kun je niet weten of het A is dan wel B die een \hat{p} vond die verder dan 3,1 % van p ligt. Anderzijds is het zo dat beide betrouwbaarheidsintervallen een doorsnede hebben, zodat ze beide de gezochte populatieproportie kunnen bevatten, al is dit niet gegarandeerd zo.

- 2 Zelfde vraag indien B een steekproefproportie van 51,3 % vond.

In dit geval is er geen overlap meer van beide betrouwbaarheidsintervallen, maar het blijft onmogelijk om te weten of A, B of beide een interval vonden dat p niet bevat.

Opdracht 35 bladzijde 41**Sensatiezucht of onoplettendheid?**

Er waaide nogal wat stof op toen begin 2002 in de Amerikaanse kranten verscheen dat 25 % van alle alcohol in de VS wordt geconsumeerd door jongeren van 12 tot 20 jaar. Dat bleek uit een studie van een centrum dat strijdt tegen drank- en drugsmisbruik.

De resultaten waren berekend uit een steekproef van 25 500 personen, van wie er 9759 een leeftijd tussen 12 en 20 jaar hadden. Die groep dronk ongeveer 25 % van de hoeveelheid alcohol die door de hele steekproef verbruikt werd.

- 1 Welke fout wordt hier gemaakt? Met andere woorden, waarom kun je uit de gegevens voor die steekproef niet besluiten dat jongeren van 12 tot en met 20 jaar zo'n 25 % van alle alcohol verbruiken die Amerikanen van 12 of ouder consumeren?

De groep 12- tot 20-jarigen is onevenredig sterk vertegenwoordigd in de steekproef.

(De organisatie publiceerde trouwens zelf zeer snel na het verschijnen van het rapport een rechtzetting, waarin ze bekende dat ze vergeten was een correctie uit te voeren voor deze onevenredige vertegenwoordiging.)

- 2 Is het werkelijke percentage groter of kleiner?

Kleiner: de groep 20- tot 70-jarigen is niet alleen groter, maar drinkt wellicht ook meer alcohol.

(Volgens de rechtzetting van de organisatie verbruikt de groep 12- tot 20-jarigen in de steekproef 11,4 % van alle alcohol.)

- 3 Zal een centrum dat drank- en drugsmisbruik bestrijdt het probleem liever belangrijk laten klinken of liever relativeren?

Er is altijd een mogelijkheid dat een organisatie als het *National Center on Addiction and Substance Abuse* dergelijke extreme cijfers graag gebruikt om haar acties te verantwoorden of om aandacht voor haar werkzaamheden te vragen.

Een aspect van 'statistische geletterdheid' is het kritisch onderzoeken van niet alleen de boodschap, maar ook de boodschapper. Bij dit onderzoek is de vergissing gelukkig zo flagrant, dat ze niet onopgemerkt kon blijven. Maar bij kleinere fouten kunnen op die manier foute uitspraken voor waar aanvaard worden.

Vandaar dat het belangrijk is dat een statistische studie altijd zeer zorgvuldig de gevolgen van de procedure documenteert; rapporten die dat niet doen, zijn steeds te wantrouwen.

Opdracht 36 bladzijde 42

Je hebt alle begrippen van dit hoofdstuk goed verwerkt, indien je de volgende vragen vlot en met behulp van concrete illustraties kan beantwoorden.

- 1** Omschrijf het begrip steekproefvariabiliteit.

Verschillende steekproeven, volgens hetzelfde steekproefontwerp samengesteld, uit eenzelfde populatie leveren verschillende steekproefresultaten op.

- 2** Omschrijf het begrip steekproefverdeling en leg uit wat het nut hiervan is.

Steekproefverdeling: het patroon dat de verschillende steekproefresultaten, afkomstig van een groot aantal aselechte steekproeven met zelfde grootte uit een zelfde populatie, vertonen.

De steekproefverdeling laat toe een foutenmarge en een betrouwbaarheidsinterval toe te kennen aan een individueel steekproefresultaat, bij een bepaald betrouwbaarheidsniveau.

- 3** Waarom hadden we die 1000 interviewers nodig, die elk 1200 personen enquêteerden? Hadden we niet beter 1 grote steekproef van 1 200 000 personen genomen?

Deze vraag verradt dat de persoon die ze stelt nog niet goed begrijpt wat het doel is van de simulatie. Ze heeft immers *niet* tot doel een schatting te geven van de onbekende populatieproportie p , maar wil de *variabiliteit* in kaart brengen en dat vereist dat de procedure heel vaak herhaald wordt.

Voor het schatten van een populatieproportie p is het uiteraard interessanter om een steekproefgrootte van 1 200 000 te hebben dan van 1200: hoe groter de steekproef, hoe kleiner de variabiliteit, m.a.w. hoe beter de steekproefproportie \hat{p} de onbekende proportie p zal benaderen, althans voor een EAS.

- 4** In de tabel op blz. 24 staat in de vierde rij, tweede kolom het getal 6,8. Wat betekent dat getal eigenlijk?

Het betekent dat, indien men heel veel steekproeven van grootte 200 zou nemen, uit een populatie met p ongeveer 40 %, men in 95 % van de gevallen een steekproefproportie \hat{p} zou vinden die niet meer dan 6,8 % afwijkt van p .

- 5** Hoe verandert de foutenmarge wanneer je het betrouwbaarheidsniveau laat afnemen? Leg uit.

Wanneer het betrouwbaarheidsniveau afneemt, wil dat zeggen dat meer betrouwbaarheidsintervallen de populatieproportie p niet zouden bevatten, mocht men heel veel steekproeven nemen. Dit doet zich voor wanneer met een kleinere foutenmarge wordt gewerkt.

