

## Antwoorden opgaven

---

### Antwoord opgave 2.6.

- a. Als je hiervoor kiest, dan kun je ook vragen stellen aan mensen die nog nooit naar het zwembad zijn geweest. Waarom niet? Kenden ze het zwembad misschien niet?
- b. Zie a. Daarbij is het goed een ondergrens aan te houden. Kleine kinderen kunnen de vragen waarschijnlijk niet beantwoorden.
- c. Zo onderzoek je de mensen die nog niet zo lang geleden in het zwembad zijn geweest en dus ook kunnen uitleggen wat er goed of niet goed aan is. Misschien moet er nog een leeftijdsgrens bij.
- d. Je kunt je afvragen of iemand nog iets zinnig kan zeggen over het zwembad als je er heel lang geleden voor het laatst bent geweest.
- e. Dit is niet zo'n goed idee. Zo vraag je alleen de mening van de zwemmers met een jaarkaart. Dat is misschien wel een heel speciale groep mensen.
- f. Bij deze doelpopulatie horen ook mensen die niet in de gemeente wonen. Dat zijn misschien wel mensen uit de regio die ook iets zinnig kunnen zeggen over het zwembad.
- g. Als de gemeente alleen maar de mening van de eigen inwoners wil weten, is dit niet zo'n goed idee. Zie ook e. Nu zitten ook jaarkaarthouders uit andere gemeenten in de doelpopulatie

### Antwoord opgave 2.7.

Antwoord a is goed. Er is sprake van onderdekking, aangezien geen enkel huishouden uit de doelpopulatie in het steekproefkader zit. De hele doelpopulatie ontbreekt.

Er is ook sprake van overdekking. In het steekproefkader zitten immers alleen maar huishoudens die niet tot de doelpopulatie behoren.

### Antwoord opgave 2.8.

- a. Deze vraag doet een beroep op de herinnering. Mensen zijn niet goed in het onthouden van (vaak onbelangrijke) zaken.
- b. De vraag bevat een dubbele ontkenning. Als je vindt dat jongeren onder de 18 alcohol moeten kunnen drinken, hoe moet je dan de vraag beantwoorden?
- c. Er zijn misschien wel twee dingen mis met deze vraag. In de eerste plaats bestaat de vraag uit twee vragen: "Heb je wel eens een coffeeshop bezocht?" en "Heb je wel eens softdrugs in een coffeeshop gekocht?". Je kunt echter maar één antwoord geven. In de tweede plaats is dit misschien wel een gevoelige vraag waarop niet iedereen een eerlijk antwoord wil geven.
- d. Dit is een suggestieve vraag. Met woorden als "toch" en "ook" wordt de suggestie gewekt dat het wel erg stom is om geen smartphone te hebben.
- e. Het is de vraag of die vraag wel te beantwoorden is. Wat wordt bedoeld met het "geheugen" van je computer? Is dat, bijvoorbeeld, de omvang van de harde schijf, of de omvang van het interne geheugen? En in welke eenheid moet je dat opgeven? Weten mensen dit eigenlijk wel?
- f. Dit is vast een vraag waarop niet iedereen een antwoord wil geven. En wat wordt bedoeld met "vaak"? Is dat elke dag, of elke week, of elke maand? Iedereen kan een ander idee hebben bij "vaak".

**Antwoord opgave 3.2.1.**

16	08	21	06	16	18	09	24	15	20	04			
----	----	----	----	----	----	----	----	----	----	----	--	--	--

Het getal 16 kwam twee keer voor. Om toch 10 verschillende getallen te krijgen, moet je dus nog een 11-de getal trekken.

**Antwoord opgave 3.2.2.**

$u$	$u \times N$	Afronden naar beneden	1 erbij tellen
0,879	26,37	26	27
0,026	0,78	0	1
0,391	11,73	11	12
0,208	6,24	6	7
0,926	27,78	27	28

**Antwoord opgave 4.6.**

- De insluitkans van elk adres was  $n / N = 209 / 9590 = 0,022$ .
- Als we het aantal personen op een adres aangegeven met  $A$ , dan is de insluitkans van een persoon op een adres gelijk aan:  $1 / A$ .
- De totale kans voor een persoon om in de steekproef te komen krijg je door beide bovenstaande kansen met elkaar te vermenigvuldigen. Dit geeft als trekkingskans:  $209 / (A \times 9590)$ .
- Een persoon in een 1-persoons huishouden heeft een kans van  $209 / 9590 = 0,022$ .
- Een persoon in een 2-persoons huishouden heeft een kans van  $209 / 19180 = 0,011$  (tweemaal zo klein).
- Personen uit kleine huishoudens zijn oververtegenwoordigd in de steekproef, en personen uit grote huishoudens zijn ondervertegenwoordigd.

**Antwoord opgave 5.2.1.**

- Schatting:  $27566 / 20 = 1378$ .
- Steekproefvariantie:  $26601160 / 19 = 1400061$ .
- Geschatte variantie van de schatter:
- $(1 - 20/341) \times 1400061 / 20 = 65897$ .
- Geschatte standaardfout van de schatter:  $\sqrt{65897} = 257$ .
- Geschatte onzekerheidsmarge:  $1,96 \times 257 = 504$ .
- Ondergrens betrouwbaarheidsinterval:  $1379 - 504 = 875$ .
- Bovengrens betrouwbaarheidsinterval:  $1379 + 504 = 1883$ .

**Antwoord opgave 5.3.1.**

- Het percentage werkenden is  $100 \times 30 / 100 = 30\%$ .
- De steekproeffractie is gelijk aan  $f = 100 / 1000 = 0,1$ . Invullen van  $n = 100$ ,  $p = 30$  en  $f = 0,1$  in formule (5.3.4) geeft je een schatting voor de variantie van het steekproefgemiddelde: 19,09. De standaardfout kun je dan schatten met de wortel hieruit, en dat is 4,37.
- Het 95%-betrouwbaarheidsinterval krijg je door  $1,96 \times 4,37$  op te tellen bij, dan wel af te trekken van 30. De ondergrens van het interval wordt dan 21,4 en de bovengrens wordt 38,6.
- Met een betrouwbaarheid van 95% kunnen we stellen dat het percentage werkenden in de populatie hoogstwaarschijnlijk zal liggen tussen 21,4% en de 38,6%.

**Antwoord opgave 5.5.**

- a. Voor het trekken van de steekproef hebben we gebruik gemaakt van de tabel met aselechte getallen. De tabel is rij voor rij afgewerkt. Van elk groepje van vijf cijfers zijn de eerste drie cijfers genomen. Zo'n drietal cijfers levert precies een getal op in de reeks van 1 t/m 1.000 (waarbij je 000 opvat als 1000). De steekproef ziet er dan zo uit:

008	631	040	293	687	342	418	948	115	076
526	692	797	197	620	812	112	421	700	783
743	100	625	833	586	410	098	277	129	930
561	395	933	871	184	516	712	491	033	454
523	626	336	268	172	491	838	238	120	765

- b. Zie de tabel hierboven.
- c. Er zijn 30 gekleurde vakjes. Het percentage mensen in de steekproef dat luistert naar de lokale omroep is dus  $100 \times 30 / 50 = 60\%$ .

- d. De berekening van de onzekerheidsmarge:

$$p = 60$$

$$\text{Variantie: } v(p) = \frac{1-f}{n-1} p(100-p) = \frac{1-50/1000}{50-1} 60(100-60) = 46,5.$$

$$\text{Standaardfout: } s(p) = \sqrt{v(p)} = \sqrt{46,5} = 6,8.$$

$$\text{Onzekerheidsmarge: } m(p) = 1,96 \times s(p) = 1,96 \times 6,8 = 13,3.$$

De onzekerheidsmarge is 13,3%. Het werkelijke percentage in de doelpopulatie kan dus 13,3 procentpunten afwijken van de schatting van 60%.

- e. Ondergrens betrouwbaarheidsinterval:  $60 - 13,3 = 46,7\%$   
 Bovengrens betrouwbaarheidsinterval:  $60 + 13,3 = 73,3\%$
- f. De schatting op basis van de peiling (60%) ligt dicht bij de werkelijke waarde (57%) in de buurt. Het was dus mogelijk een goede schatting te maken met een steekproef van 50 personen. De onzekerheidsmarge is wel erg groot. Dat komt omdat de steekproef maar klein was.

**Antwoord opgave 6.5.**

- a. Hieronder de tabel met de gegevens:

	631	040		687		418		115	
	692	797		620		112			783
			833	586			277		930
561		933	871					033	
		336	268	172		838			

- b. Zie tabel.
- c. 22 van de 50 personen hebben meegedaan aan de peiling. Het percentage respons is dus  $100 \times 22 / 50 = 44\%$ . Dat is niet zo'n hoog percentage respons. Je loopt dus risico op ernstige vertekeningen.

d. Van de 22 personen die hebben meegedaan aan de peiling hebben 14 een rood vakje. Dat zijn de luisteraars. Het percentage luisteraars is (afgerond)  $100 \times 14 / 22 = 64\%$ . Deze schatting is hoger dan de schatting in opgave 6.1 (60%). Kennelijk zijn het vooral de luisteraars die meedoen aan de peiling.

e. De berekening van de onzekerheidsmarge:

$$p = 64$$

$$\text{Variantie: } v(p) = \frac{1-f}{n-1} p(100-p) = \frac{1-22/1000}{22-1} 64(100-64) = 107,3.$$

$$\text{Standaardfout: } s(p) = \sqrt{v(p)} = \sqrt{107,3} = 10,4.$$

$$\text{Onzekerheidsmarge: } m(p) = 1,96 \times s(p) = 1,96 \times 10,4 = 20,3.$$

De onzekerheidsmarge is 20,3%. Het werkelijke percentage in de doelpopulatie kan dus 20,3 procentpunten afwijken van de schatting van 64%. De onzekerheidsmarge is veel groter dan in opgave 5.5. Dat komt omdat er maar zo weinig mensen meededen aan de peiling.

f. Ondergrens betrouwbaarheidsinterval:  $64 - 20,3 = 43,7\%$

$$\text{Bovengrens betrouwbaarheidsinterval: } 64 + 20,3 = 84,3\%$$

Het werkelijke percentage luisteraars ligt met grote waarschijnlijkheid tussen de 43,7% en 84,3%. Die conclusie kun je alleen trekken als de non-respons geen vertekening veroorzaakt. Helaas is dit in de praktijk meestal wel het geval.

g. Door het optreden van non-respons is de schatting gaan schuiven van 60% naar 64%. Dat is verder weg van werkelijke waarde in de populatie (57%). Dat komt omdat de non-respons zeer waarschijnlijk selectief is en dus een vertekening veroorzaakt.

### Antwoord afsluitende opgave 8.1.

a. Het steekproefgemiddelde is gelijk aan:  $\bar{y} = 1555 / 20 = 77,75$ .

De geschatte populatievariantie is:  $s^2 = 122,6184$ .

De schatting voor de variantie van het steekproefgemiddelde luidt:

$$v(\bar{y}) = (1/n - 1/N) \times s^2 = (1/20 - 1/10\,000) \times 122,6184 = 6,118659$$

De schatting voor de standaardfout (de standaardafwijking van het steekproefgemiddelde) is:

$$s(\bar{y}) = \sqrt{6,118659} = 2,4736$$

b. Het 95%-betrouwbaarheidsinterval is:

$$(77,75 - 1,96 \times 2,4736 ; 77,75 + 1,96 \times 2,4736) = (72,9 ; 82,6)$$

Deze uitkomst sluit niet uit dat de gemiddelde tevredenheidsscore 80 of hoger zou kunnen zijn, want 80 ligt in het betrouwbaarheidsinterval.

c. Toepassen van formule (5.4.8) geeft:

$$n \geq \frac{1}{\left(\frac{Max}{1,96s}\right)^2 + \frac{1}{N}} = \frac{1}{\left(\frac{2}{1,96 \times 11,07332}\right)^2 + \frac{1}{10\,000}} = 116,39206$$

Dus de steekproefomvang  $n$  moet minimaal gelijk zijn aan 117.

### Antwoord afsluitende opgave 8.2.

a. Percentage voorstanders is  $100 \times (128 + 60) / (128 + 60 + 512 + 40) = 100 \times 188 / 740 = 25,4\%$   
Percentage non-respons is  $100 \times 260 / 1000 = 26 \%$

b. Je krijgt de ondergrens door aan te nemen dat alle non-respondenten tegenstanders zijn:  
 $100 \times (188 / 1000) = 18,8 \%$

Je krijgt de bovengrens door aan te nemen dat alle non-respondenten voorstanders zijn:  
 $100 \times ((188 + 260) / 1000) = 100 \times 448 / 1000 = 44,8 \%$

Het werkelijke percentage voorstanders in de gehele steekproef zal dus liggen tussen de 18,8% en de 44,8%. Dat is een flinke bandbreedte.

c. Het percentage autobezitters in de respons is  $100 \times 640 / 740 = 86,486\%$ .  
Het percentage autobezitters in de doelpopulatie is 80,000%.  
Het gewicht voor de autobezitters is  $80,000 / 86,486 = 0,925$ .

Het percentage niet-autobezitters in de respons is  $100 \times 100 / 740 = 13,514 \%$

Het percentage niet-autobezitters in de doelpopulatie is 20,000 %.

Het gewicht voor de niet-autobezitters is  $20,000 / 13,514 = 1,480$ .

d. De tabel met gewogen aantallen:

	Voor rekeningrijden	Tegen rekeningrijden
Heeft auto	$128 \times 0,925 = 118,4$	$512 \times 0,925 = 473,6$
Heeft geen auto	$60 \times 1,480 = 88,8$	$40 \times 1,480 = 59,2$

e. Het gewogen percentage voorstanders is gelijk aan:

$100 \times (118,4 + 88,8) / (118,4 + 88,8 + 473,6 + 59,2) = 100 \times 207,2 / 740 = 28\%$ .

Het percentage voorstanders is groter geworden. Dat komt doordat de voorstanders als gevolg van non-respons ondervertegenwoordigd waren in de ongewogen steekproef. Het waren vooral tegenstanders die meededen aan de peiling.