

## 6.3 Verschil kwalitatieve variabelen

### Inleiding

Kun je zonder meer zeggen dat mannen langer zijn dan vrouwen? Dat er meer meisjes wiskunde A kiezen dan jongens? Dat Belgen meer vreemde talen spreken dan Nederlanders?

Bij de verschillende soorten statistische variabelen horen verschillende meetniveau's, die ken je al. Als je verschillen tussen statistische variabelen in kaart wil brengen, hangt de manier waarop je dit kunt doen af van het meetniveau van deze variabelen.

#### Je leert in dit onderwerp

- het verschil tussen twee kwalitatieve variabelen te beschrijven en te interpreteren;
- te werken met vuistregels om deze verschillen te beschrijven.

#### Voorkennis

- soorten statistische variabelen herkennen;
- de begrippen onderzoek, steekproef, populatie en representatief.

### Verkennen

#### Opgave V1

Langzamerhand kom je in de buurt van echt statistisch onderzoek. Hier een klein voorproefje.

Probeer te beschrijven hoe je zou moeten onderzoeken of meisjes naar verhouding structureel vaker wiskunde A kiezen en jongens naar verhouding wiskunde B.

#### Uitleg 1

Bekijk de tabel. In deze kruistabel is de waardering, de voorstelingsbeleving, gegeven van een groep leerlingen die een toneelvoorstelling heeft bezocht, uitgesplitst per examenprofiel.

Beide variabelen in deze kruistabel zijn kwalitatief. Variabele voorstelingsbeleving heeft oplopende waarden en heet daarom een ordinale variabele.

Op grond van deze tabel lijkt het alsof de EM-leerlingen de voorstelling hoger hebben gewaardeerd dan de NG-leerlingen, maar die groepen zijn niet even groot. Om goed te kunnen vergelijken bij groepen die niet even groot zijn, moet je de absolute frequenties eerst omzetten in relatieve frequenties.

Bekijk de tabel. In deze tabel staan onder  $p$  de percentages en onder  $cp$  de cumulatieve percentages.

	CM	EM	NG	NT
1 = niet boeiend	5	8	6	17
2 = gaat wel	12	12	18	13
3 = boeiend	9	18	15	8
4 = erg boeiend	9	10	6	2

Tabel 1

	EM			NG			$V_{cp}$
	aantal	$p$	$cp$	aantal	$p$	$cp$	
1 = niet boeiend	8	16,7	16,7	6	13,3	13,3	3,3
2 = gaat wel	12	25,0	41,7	18	40,0	53,3	11,7
3 = boeiend	18	37,5	79,2	15	33,3	86,7	7,5
4 = erg boeiend	10	20,8	100,0	6	13,3	100,0	0,0
	48			45			

Tabel 2

De percentages onder  $p$  van de EM-leerlingen en de NG-leerlingen verschillen: procentueel hebben veel meer NG-leerlingen 'gaat wel' geantwoord dan dat dit bij de EM-leerlingen het geval was. En

procentueel hebben veel minder NG-leerlingen ‘erg boeiend’ geantwoord dan dat dit bij de EM-leerlingen het geval was. Toch hebben procentueel ook minder NG-leerlingen ‘niet-boeiend’ geantwoord. Een conclusie trekken is nog niet eenvoudig. Daarom is voor ordinale kwalitatieve variabelen de volgende afspraak gemaakt:

het maximale verschil van de cumulatieve percentages  $\max V_{CP}$  wordt gebruikt als maat voor het verschil tussen twee groepen.

Vuistregels voor de interpretatie van de  $\max V_{CP}$  zijn:

- Als  $\max V_{CP} \leq 20\%$  dan is het verschil ‘gering’.
- Als  $20\% < \max V_{CP} \leq 40\%$  dan is het verschil ‘middelmattig’.
- Als  $\max V_{CP} > 40\%$  dan is het verschil ‘groot’.

Hier is  $\max V_{CP} \leq 20\%$ , dus het verschil is gering.

### Opgave 1

Bekijk **Uitleg 1**.

- Leg uit waarom de variabele voorstellingsbeleving een kwalitatieve variabele is en geen kwantitatieve variabele.
- Ordinale kwalitatieve variabelen zie je vaak terug in enquêtes. Bedenk zelf een ordinale kwalitatieve variabele die in een enquête zou kunnen voorkomen en geef ook de mogelijke waarden en hun volgorde.

### Opgave 2

Bekijk **Uitleg 1**.

Vergelijk de voorstellingsbeleving van de NG-leerlingen met de NT-leerlingen. Bereken eerst de bijbehorende  $\max V_{CP}$ .

### Uitleg 2

Om te onderzoeken of vitamine C helpt tegen verkoudheid kregen 139 personen vitamine C toegediend en 140 personen een placebo (fopmiddel). De proefpersonen wisten niet in welke categorie ze zaten.

	verkouden	niet verkouden
vitamine C	19 (a)	122 (b)
placebo	29 (c)	109 (d)

Tabel 3

De onderzoeker telde het aantal personen dat verkouden werd en maakte deze tabel.

Met deze aantallen kun je de groepen vergelijken. Je gebruikt hierbij het getal *phi* ofwel  $\varphi$ :

- Noem de waarden in de vier cellen van de kruistabel *a*, *b*, *c* en *d*.
- Vul deze waarden in de formule voor  $\varphi = \varphi$  in:

$$\varphi = \frac{ad - bc}{\sqrt{(a+b) \cdot (a+c) \cdot (b+d) \cdot (c+d)}}$$

- Je vindt:  $\varphi \approx -0,100$ .

Trek conclusies over de mate van verschil tussen beide groepen met de volgende vuistregels:

- Als  $\varphi < -0,4$  of  $\varphi > 0,4$  is de conclusie: het verschil is ‘groot’.
- Als  $-0,4 \leq \varphi < -0,2$  of  $0,2 < \varphi \leq 0,4$  is de conclusie: het verschil is ‘middelmattig’.
- Als  $-0,2 \leq \varphi \leq 0,2$  is de conclusie: het verschil is ‘gering’.

Volgens deze vuistregels is in dit onderzoek sprake van een gering verschil.

### Opgave 3

Over het algemeen maak je een kruistabel om een beeld te krijgen van het verband en/of verschil tussen twee variabelen. De kruistabel in **Uitleg 2** bevat absolute frequenties en dat is de minst inzichtelijke manier om een kruistabel samen te stellen.

Verander de absolute frequenties in de kruistabel van de uitleg in relatieve frequenties nadat je eerst hebt besloten of je horizontaal percenteert, verticaal of beide (wat vind je zinvol in deze situatie?) en geef aan of je daarmee ook al iets kunt zeggen over de mate van verschil tussen de twee variabelen.

### Opgave 4

Bekijk de kruistabel in **Uitleg 2**. Er is een vaste plek in de kruistabel voor de frequenties  $a$ ,  $b$ ,  $c$  en  $d$  die gebruikt worden om verschilmaat  $\varphi$  te berekenen. Je kunt je afvragen of het dan ook van belang is hoe je de kruistabel samenstelt.

Maak rijen van de kolommen in de kruistabel en vice versa en bepaal voor deze nieuwe kruistabel de waarde van  $\varphi$ .

Is deze waarde anders dan de waarde die in de uitleg is berekend?

## Theorie en voorbeelden

### Om te onthouden

In statistisch onderzoek worden vaak twee groepen onderzocht en met elkaar vergeleken: in welke mate is er verschil? Er zijn legio methodes om inzicht in het verschil te krijgen, al naar gelang het soort kenmerk en dus het type variabele waarbij het verschil tussen de twee groepen wordt vergeleken.

Als de variabele kwalitatief is maar wel een ordening kent in de mogelijke waarden, dan spreek je van een **ordinale kwalitatieve variabele**. Bereken per groep de cumulatieve relatieve frequenties van de mogelijke waarden. Bepaal vervolgens per waarde het verschil in cumulatieve relatieve frequentie tussen de twee groepen. De waarde met het hoogste verschil heet officieel **max  $V_{cp}$** .

Vuistregels voor de interpretatie van  $\max V_{cp}$  zijn (zie ook dit **Formuleoverzicht**):

- Als  $\max V_{cp} \leq 20\%$  dan is het verschil 'gering'.
- Als  $20\% < \max V_{cp} \leq 40\%$  dan is het verschil 'middelmatic'.
- Als  $\max V_{cp} > 40\%$  dan is het verschil 'groot'.

	groep 1	groep 2
waarde 1	$a$	$b$
waarde 2	$c$	$d$

Tabel 4

Als de variabele kwalitatief is en geen ordening heeft spreek je van een **nominale kwalitatieve variabele**. Als hij slechts twee waarden heeft maak je een 2-bij-2 **kruistabel** van de absolute frequenties.

Bereken daarmee de verschilmaat **phi** ( $\varphi$ ) volgens de formule:

$$\varphi = \frac{ad-bc}{\sqrt{(a+b) \cdot (a+c) \cdot (b+d) \cdot (c+d)}}$$

Vuistregels voor de interpretatie van  $\varphi$  zijn (zie ook dit **Formuleoverzicht**):

- Als  $-0,2 \leq \varphi \leq 0,2$  dan is het verschil is 'gering'.
- Als  $-0,4 \leq \varphi < -0,2$  of  $0,2 < \varphi \leq 0,4$  dan is het verschil 'middelmatic'.
- Als  $\varphi < -0,4$  of  $\varphi > 0,4$  dan is het verschil 'groot'.

### Voorbeeld 1

Aan een aantal hotelgasten is gevraagd om het dinerbuffet te beoordelen. In de tabel staan de resultaten uitgesplitst naar werelddeel van herkomst.

	Europa	Afrika	Azië	Noord-Amerika	Zuid-Amerika	Oceanië
1 = zeer goed	12	4	6	3	4	0
2 = goed	15	6	22	6	8	4
3 = matig	10	2	15	4	2	2
4 = slecht	3	0	4	4	0	2

Tabel 5

Op grond van deze tabel lijkt het alsof de Europeanen het dinerbuffet hoger hebben gewaardeerd dan de Aziaten, maar die groepen zijn niet even groot.

Welke mate van verschil is er tussen de Europeanen en de Aziaten en hun oordeel over het dinerbuffet?

Antwoord

In deze situatie gaat het om twee groepen, Europeanen en Aziaten, en het verschil in hun oordeel over een dinerbuffet. Variabele *buffetbeoordeling* is een ordinale nominale variabele. Dat betekent dat je verschilmaat  $\max V_{cp}$  kunt gebruiken om een conclusie over het verschil tussen beide groepen te trekken.

Bekijk in de tabel onder  $p$  de percentages, onder  $cp$  de cumulatieve percentages en onder  $V_{cp}$  het verschil van  $cp$ .

	Europa			Azië			$V_{cp}$
	aantal	$p$	$cp$	aantal	$p$	$cp$	
1 = zeer goed	12	30	30	6	12,8	12,8	17,2
2 = goed	15	37,5	67,5	22	46,8	59,6	7,9
3 = matig	10	25	92,5	15	31,9	91,5	1
4 = slecht	3	7,5	100	4	8,5	100	0
	40			47			

Tabel 6

Er geldt  $\max V_{cp} = 17,2\%$  en dus  $\max V_{cp} \leq 20\%$ .

Volgens de vuistregels is het verschil gering.

### Opgave 5

Bekijk de tabel over de waardering van een dinerbuffet in [Voorbeeld 1](#).

Deze opgave betreft het verschil in beoordeling van dit dinerbuffet door enerzijds Europeanen en anderzijds Noord-Amerikanen.

- Wat is je conclusie over dat verschil als je alleen kijkt naar de absolute frequenties die in de tabel staan?
- Wat is je conclusie over dat verschil als je alleen kijkt naar de relatieve frequenties die in de tabel staan?
- Welke mate van verschil is er op statistisch verantwoorde wijze te vinden tussen Europeanen en Noord-Amerikanen en hun oordeel over het dinerbuffet?

### Opgave 6

Op een nieuwssite staat een onderzoek naar de beoordeling van een nieuwe app. Bekijk de resultaten.

	12- tot 15-jarigen	16- tot 19-jarigen
1 = leuk	31	22
2 = niet leuk	27	38

Tabel 7

- a Welke mate van verschil in beoordeling van de app zit er tussen de groep van 12- tot 15-jarigen en de groep van 16- tot 19-jarigen?
- b Was je tot dezelfde conclusie gekomen als je alleen de absolute frequenties had bekeken?
- c De mensen die het onderzoek uitgevoerd en/of gepubliceerd hebben, gaven beoordeling 'leuk' de waarde 1 en beoordeling 'niet leuk' waarde 2. Dit hadden ze ook andersom kunnen doen. Zou dat hebben uitgemaakt voor de bepaling van de mate van verschil? Licht je antwoord toe met berekeningen.

### Voorbeeld 2

Onderzoek onder een groep mannen en vrouwen naar de bijwerkingen van een medicijn gaf deze resultaten.

Bepaal de mate van verschil tussen het ontstaan van bijwerkingen bij mannen en bij vrouwen.

	bijwerking	geen bijwerking
mannen	86	14
vrouwen	61	39

Tabel 8

Antwoord

Het gaat hier om het verschil tussen een groep mannen en vrouwen. *Bijwerking* is een nominale variabele met twee mogelijke waarden: er kan een 2-bij-2 tabel van worden gemaakt en om echt een uitspraak over de mate van verschil te kunnen doen, kun je verschilmaat  $\varphi$  berekenen met behulp van de gegeven formule.

Je vindt:  $\varphi \approx 0,28$

Volgens de vuistregels voor de interpretatie van  $\varphi$  is er dus sprake van een middelmatig verschil tussen de bijwerkingen bij mannen en de bijwerkingen bij vrouwen.

### Opgave 7

Op een nieuwssite staat een onderzoekje naar de beoordeling van een nieuwe app. Bekijk de resultaten.

	12- tot 15-jarigen	16- tot 19-jarigen
1 = leuk	31	22
2 = niet leuk	27	38

Tabel 9

In **Opgave 6** heb je een verschilmaat berekend, namelijk  $\max V_{CP}$ . Hiermee bleek dat er sprake was van een gering verschil tussen de beide groepen omtrent hun beoordeling van de app.

- a Kijk nu met behulp van verschilmaat  $\varphi$  in welke mate er verschil is tussen deze groepen en vergelijk de conclusie op basis van  $\varphi$  met de conclusie op basis van  $\max V_{CP}$ .
- b Leg uit waarom je voor deze gegevens zowel  $\max V_{CP}$  als  $\varphi$  kunt gebruiken als verschilmaat. Zou dit ook kunnen met de gegevens uit **Voorbeeld 2**?

### Opgave 8

Bekijk de gegevens in de tabel.

- a Bereken verschilmaat  $\varphi$  voor deze gegevens en trek daarmee een conclusie over het verschil tussen de twee groepen.
- b Laat zien dat je bij deze gegevens helemaal geen verschilmaat hoeft te gebruiken: je kunt namelijk met behulp van de absolute frequenties (of met de relatieve frequenties, als je het niet zo snel ziet) al laten zien dat er helemaal geen verschil is tussen de twee groepen.

	groep 1	groep 2
waarde 1	45	62
waarde 2	135	186

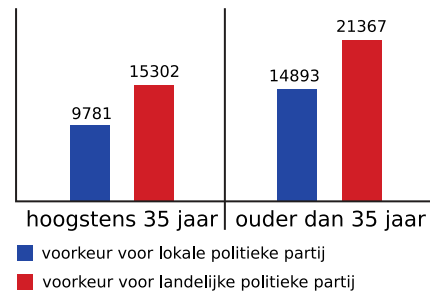
Tabel 10

## Verwerken

### Opgave 9

Bekijk de figuur. Je ziet de voorkeuren voor lokale en landelijke politieke partijen in een Nederlandse gemeente afhankelijk van de leeftijdsgroep.

- a Maak een kruistabel voor deze gegevens.
- b Leg uit waarom je de mate van verschil tussen het stemgedrag van stemgerechtigde inwoners die hoogstens 35 jaar zijn en degenen die ouder zijn dan 35 jaar in ieder geval met verschilmaat  $\varphi$  kunt bepalen.



Figuur 1

De waarde van  $\varphi$  voor deze gegevens is -0,02.

- c Bepaal op basis van deze waarde met de vuistregels de mate van verschil tussen het stemgedrag van stemgerechtigde inwoners die hoogstens 35 jaar zijn en degenen die ouder zijn dan 35 jaar.

### Opgave 10

Bekijk de drie 2-bij-2 kruistabellen met absolute frequenties die verzameld zijn voor een statistisch onderzoek.

	school A	school B
jonge docenten	83	78
oude docenten	115	47

kruistabel 1

	workshop tekenen	workshop muziek
klas A	20	10
klas B	14	7

kruistabel 2

uurloon voor vakkenvuller (€)	supermarkt A	supermarkt B
2,5– < 3,0	12	19
3,0– < 3,5	7	15

kruistabel 3

Tabel 11

Beantwoord de vragen voor ieder van de drie kruistabellen.

- a Geef bij elke tabel aan welk verschil tussen de twee groepen er waarschijnlijk wordt bestudeerd. Geef ook het type variabele dat onderzocht wordt.
- b Heeft de onderzoeker bij deze gegevens verschilmaat  $\max V_{CP}$  en/of  $\varphi$  nodig voor een conclusie over het verschil tussen de twee groepen of volstaat hier het vergelijken van percentages? Licht je antwoord toe.

- c Als de onderzoeker verschilmaat  $\max V_{CP}$  of  $\varphi$  wil gebruiken om de mate van verschil tussen de groepen te bepalen, welke komt/komen dan in aanmerking voor deze gegevens? Licht je antwoord toe.

### Opgave 11

Bekijk de tabel met de relatieve frequenties van het hoogst genoten onderwijs in twee regio's in een land.

	regio west	regio oost
basisschool	6	3
lager voortgezet onderwijs	11	19,5
hoger voortgezet onderwijs	5,5	8
MBO	40,5	37
HBO	28	20,5
WO	9	12

Tabel 12

Welke mate van verschil in hoogst genoten onderwijs bestaat er tussen de beide regio's?

### Opgave 12

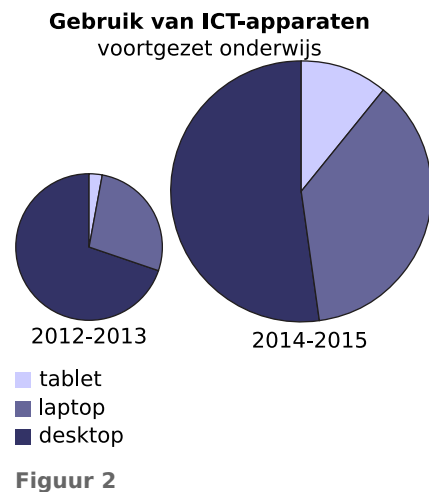
In een oud rapport van Kennisnet over ICT-gebruik in het onderwijs stonden deze gegevens over het gebruik van ICT-apparaten op scholen in het voortgezet onderwijs.

Leg uit in hoeverre je een betrouwbare uitspraak kunt doen omtrent het verschil in gebruik van ICT-apparaten tussen schooljaar 2012-2013 en schooljaar 2014-2015 op basis van de in a t/m d genoemde gegevens.

Geef aan welk verschil je ermee kunt onderzoeken (als het kan) en hoe betrouwbaar de conclusie is die je eruit kunt trekken.

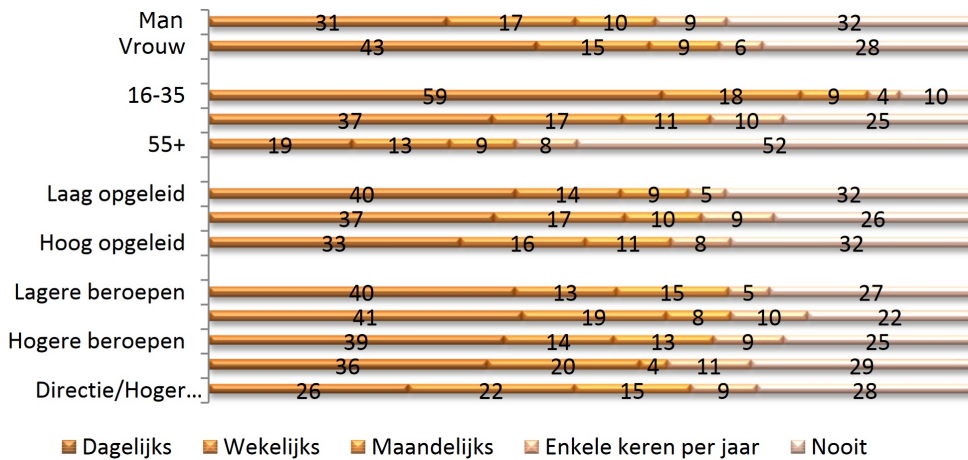
- a De vorm van de cirkeldiagrammen.
- b De percentages die bij iedere cirkelsector horen.
- c  $\max V_{CP}$
- d  $\varphi$

(bron: Kennisnet)



### Opgave 13

Bekijk de figuur. In een rapport van de Universiteit Twente staat een overzicht over het bezoek aan sociale netwerksites.



Figuur 3

Onderzoek de mate van verschil tussen mannen en vrouwen wat betreft de frequentie van hun bezoek aan sociale netwerksites.

### Opgave 14

Een farmaceutisch bedrijf heeft een studie gedaan naar het ontstaan van hartfalen tijdens het slikken van hun nieuwe combinatietablet met aspirine. Een journalist is erg geïnteresseerd in de resultaten van deze studie, maar het bedrijf is niet erg scheutig met informatie. Hij weet ondertussen wel het volgende:

- de steekproefgrootte van de studie is 22071;
- het totale aantal personen in de steekproef dat géén last van hartfalen kreeg tijdens de studie is 21693;
- het aantal personen in de studie dat aspirine slikte en géén last van hartfalen kreeg tijdens de studie is 10898;
- het aantal personen in de studie dat géén aspirine slikte maar wel last van hartfalen kreeg tijdens de studie is 239.

Volgens de journalist heeft hij genoeg aan deze gegevens om zelf de mate van verschil te berekenen tussen het ontstaan van hartfalen tussen de groep die wel aspirine slikte en de groep die dat niet deed.

- a Laat zien dat dat inderdaad kan en trek een conclusie omtrent de mate van het genoemde verschil. Stel je de situatie voor dat de journalist alleen de absolute frequenties te horen kreeg zoals getoond in deze tabel:

	hartfalen	geen hartfalen
aspirinegebruikers		10898
geen aspirinegebruikers	239	10795

Tabel 13

Hij kan dan de minimale hoeveelheid personen berekenen die aspirine slikten en tijdens de studie hartfalen kregen zodanig dat het verschil tussen de groep die wel aspirine slikte en de groep die dat niet deed 'groot' te noemen is. Met behulp van deze kennis kan hij het farmaceutische bedrijf streng ondervragen.

- b Stap nogmaals in de schoenen van de journalist en bereken dit minimumaantal.



## Toepassen

### Opgave 15: Intelligentie van honden

Voor een onderzoek naar de intelligentie van hondenrassen is aan hondenbezitters gevraagd hoelang het duurde voor de hond naar zijn naam luisterde vanaf het moment dat ze hem als pup kregen. De resultaten van labradors en beagles staan in de tabel.

	labrador	beagle
luistert naar naam binnen een week	55	39
luistert naar naam na meer dan een week	32	27

Tabel 14

Bepaal op basis van deze tabel de mate van verschil in intelligentie tussen een labrador en een beagle.

### Opgave 16: Geen verschil tussen beide groepen

Een onderzoek spitst zich toe op de vraag of er verschil in  $Y$ -waarde bestaat tussen de groepen met  $X_1$  en met  $X_2$ .

Het resultaat halverwege het onderzoek naar eigenschappen  $X$  en  $Y$  in een populatie van 850 elementen zie je in de tabel.

Onderzoek hoe de kruistabel ingevuld moet zijn als er geen enkel verschil zit tussen beide groepen.

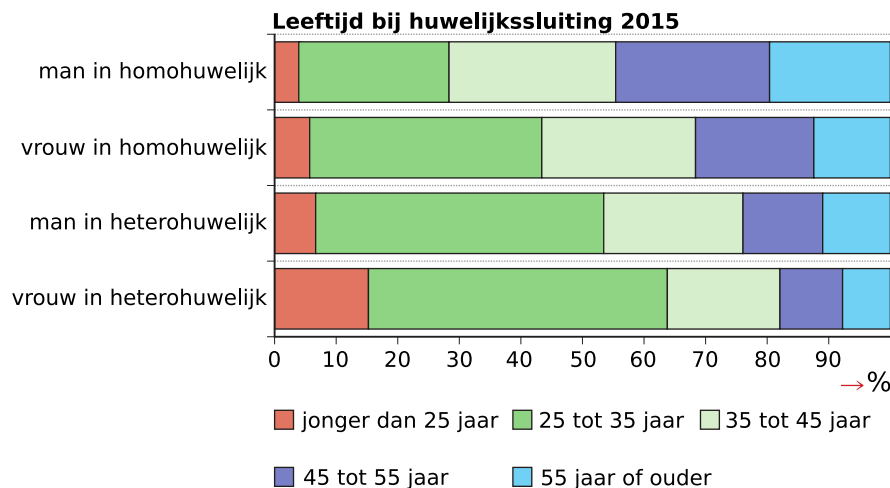
	$X_1$	$X_2$	
$Y_1$	...	...	102
$Y_2$	...	...	
		382,5	850

Tabel 15

## Testen

### Opgave 17

Bekijk de leeftijd bij huwelijkssluiting in 2015, gegroepeerd naar soort huwelijk.



Figuur 4

Maak een cumulatieve tabel en trek op grond van deze tabel de conclusies van de soort "Er is een middelmatig verschil tussen het aantal ..." voor dezelfde leeftijdscategorieën.

**Opgave 18**

Dit is het resultaat van een statistisch onderzoek.

	huid verbrandt snel	huid verbrandt niet snel
rood haar	55	8
bruin haar	87	102


**Tabel 16**

- a** Trek op grond van deze gegevens een conclusie over het verschil tussen roodharigen en bruinharigen met betrekking tot het verbranden van de huid door de zon. Gebruik daarbij het getal  $\varphi$  en de bijbehorende vuistregels.
- b** Trek op grond van percentages een conclusie over het verschil tussen roodharigen en bruinharigen met betrekking tot het verbranden van de huid door de zon. Gebruik de vuistregels:
- Als het verschil in percentage groter of gelijk is aan 40% is de conclusie:  
Het verschil is groot.
  - Als het verschil in percentage tussen 20% en 40% is, is de conclusie:  
Het verschil is middelmatig.
  - Als het verschil in percentage kleiner of gelijk is aan 20% is de conclusie:  
Het verschil is gering.



© 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.

---

