

# 3. Data verwerven

Boekje 3 havo wiskunde A, domein E: Statistiek

Uitwerkingen



Verantwoording



© 2015, SLO (nationaal expertisecentrum leerplanontwikkeling), Enschede

Dit lesmateriaal is ontwikkeld in het kader van de nieuwe examenprogramma's zoals voorgesteld door de commissie Toekomst Wiskunde Onderwijs (cTWO) en herzien door SLO.

Mits de bron wordt vermeld, is het toegestaan zonder voorafgaande toestemming van de uitgever deze uitgave geheel of gedeeltelijk te kopiëren en/of verspreiden en om afgeleid materiaal te maken dat op deze uitgave is gebaseerd.

Auteurs: Erik van Barneveld, Wouter Boer, Carel van de Giessen, Peter Kop, Heleen van der Ree, Henk Reuling, Frits Spijkers, Tanja Stroosma, Anneke Verschut

Met medewerking van: Nico Alink, Martine de Klein (eindredactie)

**Informatie:** SLO  
Afdeling: tweede fase  
Postbus 2041, 7500 CA Enschede  
Telefoon (053) 4840 661  
Internet: [www.slo.nl](http://www.slo.nl)  
E-mail: [tweedefase@slo.nl](mailto:tweedefase@slo.nl)



## Inhoud

§ 3.1	Onderzoeks- en enquêtevragen .....	4
§ 3.2	Steekproeven en fouten .....	6
§ 3.3	Standaardafwijking .....	7
§ 3.4	Steekproeffout: variatie bij steekproeven.....	8
§ 3.5	Normale verdeling. ....	9
§ 3.6	Toevallige steekproeffouten in getallen .....	14



## § 3.1 Onderzoeks- en enquêtevragen

### Opgave 3

Mogelijke onderzoeksvragen:

- Is het percentage havoleerlingen t.o.v. vwo-leerlingen dat naar een hbo-opleiding gaat groter dan 10 jaar geleden?*
- Is het gemiddelde cijfer voor een vak waar men bijles voor gevolgd heeft hoger geworden sinds de bijles?*
- Hoeveel uren per dag besteedt een havo 5-leerling aan sport/computerspelletjes/lezen/chillen?*
- Hoeveel procent van de leerlingen die na 5 havo gaan werken zijn na 5 jaar nog tevreden met dit besluit?*
- Wat zijn voor meisjes tussen 8 en 18 jaar de belangrijkste redenen voor een vriendschap met een ander meisje?*
- Hoeveel onderwerpen uit de politiek van het afgelopen jaar kunnen havo 5-leerlingen zich herinneren?*
- Hoeveel lessen hebben de havo 5-leerlingen afgelopen jaar gespijeld?*

### Opgave 4

Mogelijke onderzoeksvragen:

- *Welke smartphone gebruiken 15- en 16-jarigen in Nederland het meest in schooljaar 2015/2016?*
- *Op welke politieke partij stemmen de meeste 15- en 16-jarigen in Nederland in 2015?*
- *Welke sport beoefenen 15- en 16-jarigen in Nederland het minst in schooljaar 2015/2016?*

### Opgave 5

Vragen b t/m f en vraag h meten allen ongeveer hetzelfde.

### Opgave 6

- De vraag wekt de suggestie dat je hem wel gezien moet hebben.  
Beter is: *Ken je de film GREASE? Heb je hem zelf gezien?*
- De vraag geeft geen specificering van wat onder boeken verstaan wordt.  
Beter is: *Hoeveel boeken heb je het afgelopen jaar voor de lol gelezen?*
- De vraag geeft de suggestie van twee kranten mee, dit kan duiden op alleen landelijke dagbladen, maar er zijn ook regionale dagbladen. Beter is: *Heb je een abonnement op een dagblad?*
- De vraag geeft geen specificatie waarom je bij de dokter geweest bent, je kunt ook met iemand mee zijn geweest. Beter is: *Hoe vaak ben je vorig jaar voor jezelf naar de dokter geweest?*
- Het is hier onduidelijk naar welk soort criminaliteit de onderzoeker vraagt. Beter is (dichter bij huis): *Denk je dat de criminaliteit in jouw wijk beter bestreden kan worden? Ja/nee.*
- Hier staat alleen een stelling. Beter is: *Wat vinden havo 5-leerlingen in Amsterdam van hun school?*
- Hier wordt naar de mening van de ouders gevraagd via de kinderen. Beter is: *Heb je het gevoel dat je ouders het belangrijk vinden dat jij betere resultaten op school haalt dan zij gedaan hebben?*
- Er wordt geen onderscheid gemaakt in wat voor soort drugs.  
Beter is: *Vind je dat iedereen in zijn eigen huis zelf mag weten of hij harddrugs gebruikt?*
- Het woord 'altijd' geeft wel iets heel definitiefs.  
Beter is: *Hoeveel glazen alcohol drink je gemiddeld voordat je uitgaat?*
- De vraag is erg suggestief.  
Beter is: *Vanaf welke leeftijd vind jij dat alcohol drinken toegestaan zou moeten worden?*
- De vraag is erg suggestief. Beter is: *Wat vind jij het maximaal aantal uren dat een havo 5-leerling per week aan huiswerk mag opkrijgen?*

- l. De vraag geeft weinig informatie. Beter is: *Wat denk je dat je volgend jaar – na je diploma – gaat studeren?*
- m. De vraag is tweeledig, dat is lastig te beantwoorden. Beter is: *Als je het voor het zeggen had en je moet een flinke som geld verdelen in het investeren in het openbaar vervoer en in snelwegen. Hoeveel procent van dat geld stop jij dan in het openbaar vervoer en hoeveel procent in snelwegen?*
- n. De vraag is suggestief.  
Beter is: *Wat vind je een goede prijs voor een treinkaartje van Den Haag naar Utrecht?*

### **Opgave 7**

Persoonlijk antwoord.



## § 3.2 Steekproeven en fouten

### Opgave 8

Lang niet ieder Nederlands huishouden staat vermeld in deze bestanden. Je krijgt hiermee geen aselechte steekproef.

### Opgave 9

Vragen waar onderzoekers lang niet altijd een (eerlijk) antwoord op krijgen zijn bijvoorbeeld:

- *Hoeveel glazen alcohol drinkt u?*
- *Hoeveel drugs gebruikt u?*
- *Hoe vaak hebt u geslachtsgemeenschap met iemand anders dan uw vaste partner?*
- *Hebt u wel eens een geslachtsziekte gehad?*

De fouten die hierdoor in je onderzoek zitten noemen we **respondent bias**.

Beter is om deze vragen in te leiden en er langzaam naar toe te werken. De onderzoeker moet aantonen dat hij betrouwbaar is, dat de antwoorden in het onderzoek absoluut anoniem blijven en dat het van groot belang is dat de geïnterviewde eerlijk antwoordt.

### Opgave 10

Zelf uit te voeren.

### Opgave 11

Het makkelijkste te voorkomen: wel in de steekproef, niet in de populatie.

*Voorbeeld van wel in steekproef, niet in populatie:*

Je wilt gegevens over een huishouden, je belt op, maar de oppas neemt de telefoon op en beantwoordt je vragen.

*Voorbeeld van niet in steekproef, wel in populatie:*

Je wilt onderzoek doen naar de verschillen tussen havo 5-leerlingen in de Randstad en in de Achterhoek, maar je vergeet om ook scholen in Rotterdam mee te nemen in je onderzoek.

### Opgave 12

Bijv. via een enquête onder jongeren die verspreid is op onderwijsinstanties en in het uitgaansleven.

*Of wellicht uit abonnementsgegevens die providers aan het CBS hebben geleverd.*

### Opgave 13

Onderzoeksvraag: *Wat vinden de leerlingen van de sfeer en het leerklimaat op school?*

Vragen:

- *Ga je met plezier naar school?*
- *Hoe voel je je op school?*
- *Stimuleert de school jou om goed je best te doen?*
- *Hoe zou je de sfeer in de klas willen omschrijven?*

Je kunt bijvoorbeeld uit iedere klas in de school random 5 leerlingen de vragen laten beantwoorden.



## § 3.3 Standaardafwijking

### Opgave 14

- Gebruik VuStat of je grafische rekenmachine of een ander hulpmiddel (vraag je docent. Dan zie je:  
Leerling A: gem. = 8,0 en s.d. = 0,46.  
Leerling B: gem. = 7,4 en s.d. = 0,86.  
Leerling C: gem. = 8,9 en s.d. = 0,19.  
Gevraagde ordening: C, A, B.
- D, B, C, A. Dit leid je als volgt af:  
De spreidingsbreedte van D en B is even kort (8), maar bij D zitten de tussenliggende waarde dichter op elkaar dan bij B.  
De spreidingsbreedte van A is net eentje langer dan van C, terwijl ze een redelijk gelijke mate van spreiding daar tussenin hebben: C zal een net iets kleinere s.d. hebben dan A.
- V, m (v is vrijwel symmetrisch verdeeld, terwijl m meer tellingen in de linker staart heeft).
- Meisjes, jongens (de cum. freq. polygoon van de meisjes is steiler).
- Jongens, meisjes (spreidingsbreedte is korter en spreiding is meer gecentreerd).
- Meisjes, jongens (verdeling heeft echt een piek).
- Dit is op zich bijna niet af te leiden, dus gebruik VuStat of je grafische rekenmachine of een ander hulpmiddel (vraag je docent). Dan zie je: meisjes (s.d. = 7,19), jongens (s.d. = 7,73).
- Gemtemp (kleinste box en kleinste spreiding), maxtemp, mintemp (grootste box en grootste spreiding).
- Hagel, onweer, sneeuw, mist
- Juli, daar wijken de temperaturen het minste af van het gemiddelde; er is weinig spreiding.
- SE, CE (dit kun je al afleiden uit de kleinere spreidingsbreedte van de SE-cijfers).

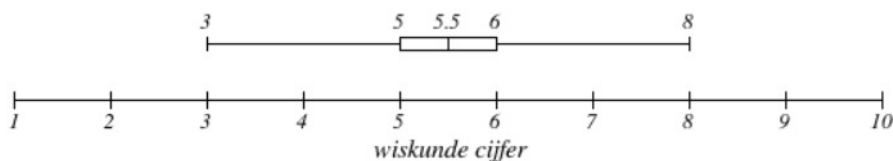
### Opgave 15

- Zie de twee staafdiagrammen voor SE en CE bij opgave 14k: CE is hier groep A en SE is groep B.
- 

Variabele A



Variabele B



- Zie de twee cum. freq. polygonen voor de lengtes van meisjes en jongens bij opgave 14d: de meisjes zijn hier groep A en de jongens zijn groep B.

### Opgave 16

Afgerond op 3 decimalen is de s.d. gelijk aan 1,732.

## § 3.4 Steekproeffout: variatie bij steekproeven

### Opgave 17

Zelf doen.

### Opgave 18

Als het goed is worden de staafdiagrammen steeds klokvormiger naarmate de hoeveelheid steekproeven toeneemt.

### Opgave 19

Zelf doen, zelf constateren.

### Opgave 20

Als je goed kijkt, zie je dat er maar één route naar de buitenste vakjes leidt. En hoe meer een vakje naar het midden toe zit, hoe meer routes er naartoe leiden. Dit is de reden waarom de verdeling er klokvormig uitziet.

### Opgave 21

Je maakt dan een plaatje van de gemiddelde jaarlijkse neerslag: eigenlijk dus een steekproefgemiddelde (en dat van heel veel jaren). In dit hoofdstuk heb je geleerd dat er dan een (symmetrische) klokvormige verdeling ontstaat. Een voorbeeld ervan zie je in paragraaf 3.5.

### Opgave 22

- a. Ja.
- b. Ja.
- c. Ja.
- d. Nee.
- e. Ja.
- f. Ja.
- g. Ja.

### Opgave 23

Zie opgave 20.





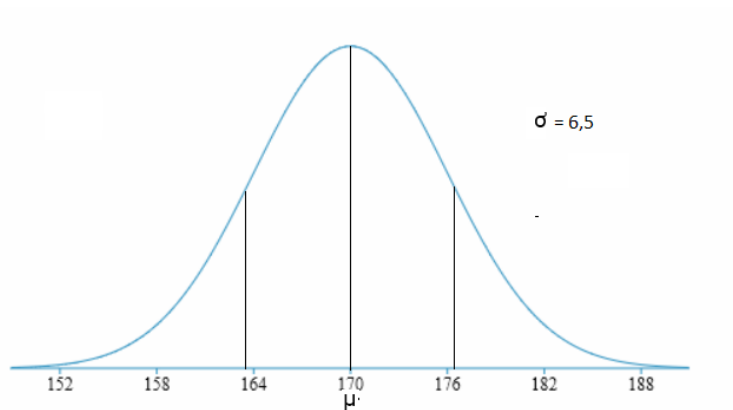
## § 3.5 Normale verdeling.

### Opgave 24

- 99,7%, dus 10,0%.
- 23,3%, dus 2,3%.
- Tussen  $\mu - \sigma$  en  $\mu + \sigma$  : 67,2%.  
Tussen  $\mu - 2\sigma$  en  $\mu + 2\sigma$  : 95,8%.

### Opgave 25

a.



- 68%.
- 2,5%.
- 97,5%.
- 0,025.

### Opgave 26

- 68%.
- 2,5%.
- 0,975.
- Als je IQ lager is dan 85.

### Opgave 27

- $\mu = 3,0$  en  $\sigma = 0,2$  percentage dat er bij hoort is 16.  
 $\mu = 82$  en  $\sigma = 6$  percentage dat er bij hoort is 84.
- Lager: de spreiding is groter, dus er zitten meer mensen verder van het gemiddelde af, daardoor is de grafiek minder hoog.

### Opgave 28

- Standaardafwijking  $\sigma$  is 15.
- Ga na.
- Klopt.



### Opgave 29

In de tabel: het gemiddelde is net iets lager dan 162 centimeter, want 53 procent van de vrouwen is kleiner of gelijk aan 162 centimeter.

15,6 procent van de vrouwen is kleiner dan 155 centimeter, de standaardafwijking is dus ongeveer 6,5 centimeter, dit klopt, ook aan de andere kant; 16 procent van de vrouwen is groter dan 168 centimeter. 95 procent van de vrouwen heeft een lengte tussen 148-175 centimeter, dit is een verschil van  $4\sigma$ , dus dat klopt ook.

### Opgave 30

Klasse	Gewicht	Percentage
I	<48 gram	2,5
II	48- 56 gram	13,5
III	56- 64 gram	34
IV	64-72 gram	34
V	72-80 gram	13,5
VI	>80 gram	2,5

### Opgave 31

Zie opgave 30.

### Opgave 32

- $\mu = 1150$  en  $\sigma = 25$ .
- 50%.
- 84%.
- De spreiding is groter.
- 2,5%.
- 2,5%.

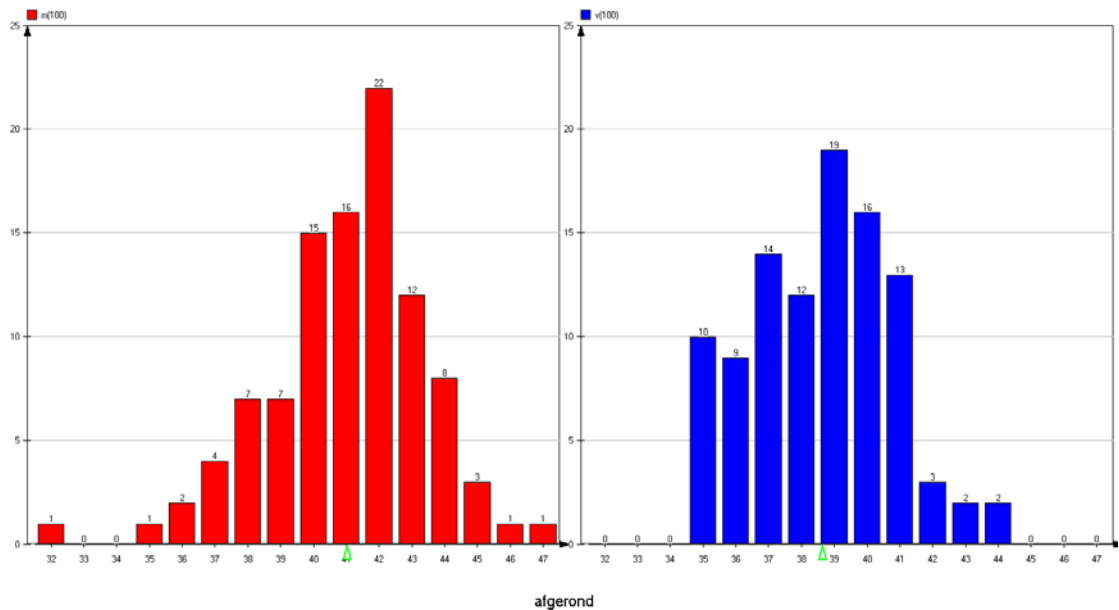
### Opgave 33

a.

Geslacht	M	V
Aantal waarnemingen	100	100
Gemiddelde	41,012	38,582
Mediaan	41,250	38,900
Modus	-	39,00
Minimum	31,60	34,60
Maximum	46,60	44,20
S.d. n-1	2,4069	2,1452
S.d. n	2,3948	2,1344
Var. n-1	5,7930	4,6017
Var. n	5,7351	4,5557

Gemiddelde voetlengte mannen = 41,0 centimeter.  
Gemiddelde voetlengte vrouwen = 38,6 centimeter.  
Standaardafwijking mannen = 2,4 centimeter.  
Standaardafwijking vrouwen = 2,1 centimeter.

b.



De benadering van de normale verdeling lijkt bij mannen iets beter dan bij vrouwen.

- c. De helft van de vrouwen heeft voetlengte 38,6 centimeter.  
93,3 procent van de mannen heeft grotere voeten.
- d.  $\mu + 2\sigma = 38,6 + 2 \times 2,1 = 42,8$  centimeter.
- e.  $\mu - 2\sigma = 41,0 - 2 \times 2,4 = 36,2$  centimeter.

**Opgave 34**

- a. Ongeveer 18%.
- b. Onder de 1750 euro.

**Opgave 35**

**Verklaring 1:**

Deze verklaring geeft niet aan waarom er zo weinig studenten in het midden zitten, bij het cijfer 5. Als het twee groepen zouden zijn, die beide normaal verdeeld zijn, dan zou de overlapping bij 5 veel groter zijn.

**Verklaring 2:**

Deze verklaring zou niet eerlijk zijn t.o.v. de studenten. Er zijn vast mensen die ongeveer in het midden zitten, deze moet je daar ook voor belonen.

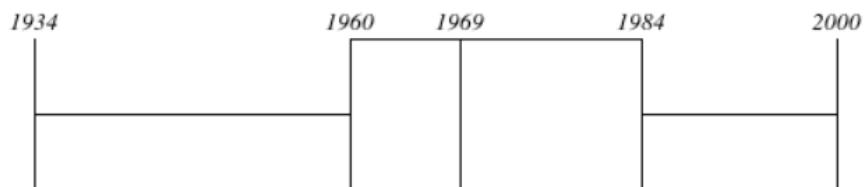
**Opgave 36**

- a. Door te kijken waar de middelste 68 procent van de waarnemingen zit.
- b. Doen.
- c.  $81,27 + 20,50 = 101,77$  dit is ongeveer 100. Dus ongeveer 84 procent heeft hoger dan 100 punten gehaald.



**Opgave 37**

- Gemiddelde geboorteaar = 1969      standaardafwijking is ongeveer 9 jaar.
- Tussen 1960 en 1969 tel je ongeveer 252 deelnemers. Dit is 32 procent. In totaal zijn er dan ongeveer 790 deelnemers.
- 



- Doen.
- Gemiddelde ongeveer 56 minuten, standaardafwijking ongeveer 5 minuten.

**Opgave 38**

- In Zweden
- Als de spreiding groter is, dan wordt de top lager. Onder iedere grafiek zit namelijk 100 procent, dus de totale oppervlakte verandert niet.

**Opgave 39**

- De top bij lengte ligt hoger, dus de spreiding is groter bij gewicht.
- Gewicht heeft een korte staart links en een langere staart rechts, bij lengte zijn de staarten ongeveer gelijk. Dit betekent dat er relatief veel mensen zijn met een hoog gewicht, terwijl erg lange mensen veel minder voorkomen.

**Opgave 40**

Minder dan 2,9 gram: 30,9%. Meer dan 2,9 gram: 69,1%.



**Opgave 41**

- a. 25,2%.
- b. 0,38%.
- c. 11,1%.
- d.  $g = 0,9953$ .

**Opgave 42**

- a. 10,9%.
- b. 51,2%.
- c. 4,1% (tussen 159,5 en 160,5 centimeter).
- d. Kleiner dan 153,7 centimeter.
- e. Langer dan 170,3 centimeter.

**Opgave 43**

- a. 10,6%, dus ongeveer 127 auto's.
- b. minder dan 48,4 seconden.
- c.  $\sigma = 2,15$  seconden.

**Opgave 44**

- a. 66,9% wordt goedgekeurd.
- b. 84,1%.
- c.  $\mu = 10,1$  en  $\sigma = 0,009$ .



## § 3.6 Toevallige steekproeffouten in getallen

### Opgave 45

Zelf doen.

### Opgave 46

- Bijvoorbeeld uit de Gemeentelijke Basis Administratie.
- 95%-gebied: tel de kleinste 125 (2,5 % van 5000) uitkomsten en de 125 grootste uitkomsten en bepaal zo de grenzen: ongeveer tussen 0,12 en 0,39.
- Zoek het linkerbuigpunt: ongeveer bij 0,18, de s.d. is dus 0,07.  
Vuistregel: 95%-gebied start bij  $0,25 - 2 * 0,07 = 0,11$  en eindigt bij  $0,25 + 2 * 0,07 = 0,39$ .
- Nu zitten alle waarnemingen grofweg tussen de 0,21 en de 0,285: de totale spreiding is dus al veel kleiner.
- Je ziet dat al aan de formule voor de steekproefproportie: de grenzen worden bepaald door (de wortel van) een breuk waarin je deelt door de steekproefomvang en als je deelt door een groter getal krijg je juist een steeds kleiner getal => de grens wordt bij een grotere steekproefomvang kleiner.

### Opgave 47

Vuistregel voor steekproefproporties geeft voor het 95%-gebied:

$$\text{als linkergrens } 0,54 - 2 \times \sqrt{\frac{0,54 \times 0,46}{1200}} \approx 0,51 \text{ en als rechtergrens } 0,54 + 2 \times \sqrt{\frac{0,54 \times 0,46}{1200}} \approx 0,57.$$

0,54 zit in dit gebied dus met 95% betrouwbaarheid kunnen we zeggen dat kandidaat A zal winnen.

### Opgave 48

Vuistregel geeft:

$$\text{linkergrens} = 0,11 - 2 \times \sqrt{\frac{0,11 \times 0,89}{1500}} \approx 0,094 \text{ en rechtergrens} = 0,11 + 2 \times \sqrt{\frac{0,11 \times 0,89}{1500}} \approx 0,126.$$

De steekproefproportie is  $\frac{136}{1500} \approx 0,091$  en dat valt NIET in het 95%-gebied.

### Opgave 49

Vuistregel voor steekproefgemiddelden geeft:

$$\text{als linkergrens } 10.000 - 2 \times \frac{2000}{\sqrt{200}} \approx 9717 \text{ en als rechtergrens } 10.000 + 2 \times \frac{2000}{\sqrt{200}} \approx 10.283 \text{ uur.}$$

### Opgave 50

Vuistregel voor steekproefproporties geeft voor het 95%-gebied:

$$\text{als linkergrens } 0,55 - 2 \times \sqrt{\frac{0,55 \times 0,45}{1988}} \approx 0,528 \text{ en als rechtergrens } 0,55 + 2 \times \sqrt{\frac{0,55 \times 0,45}{1988}} \approx 0,572.$$

61% valt hier niet binnen en past dus niet bij de te verwachten onnauwkeurigheid => het aantal is met een betrouwbaarheid van 95% toegenomen.

### Opgave 51

Vuistregel voor steekproefgemiddelden geeft:

$$\text{als linkergrens } 64 - 2 \times \frac{12}{\sqrt{50}} \approx 60,6 \text{ en als rechtergrens } 64 + 2 \times \frac{12}{\sqrt{50}} \approx 67,4 \text{ punten.}$$

Bij een steekproef van 500 kandidaten worden dat de grenzen 62,9 en 65,1 : zoals verwacht wordt het 95%-gebied smaller.

**Opgave 52**

a. Bij omvang 20 geeft de vuistregel voor steekproefproporties de volgende grenzen voor het 95%-gebied:

$$\text{linkergrens } 0,35 - 2 \times \sqrt{\frac{0,35 \times 0,65}{20}} \approx 0,137 \text{ en als rechtergrens } 0,35 + 2 \times \sqrt{\frac{0,35 \times 0,65}{20}} \approx 0,563 \Rightarrow$$

staafdiagram voldoet niet helemaal.

rest zelf doen.

b. 1 staaf of 1 hele lange staaf met twee hele kleine eromheen.

**Opgave 53**

Vuistregel voor steekproefproporties geeft voor het 95%-gebied de volgende grenzen:

$$\text{linkergrens } 0,05 - 2 \times \sqrt{\frac{0,05 \times 0,95}{50}} \approx -0,012 \text{ (0 % dus) en}$$

$$\text{rechtergrens } 0,05 + 2 \times \sqrt{\frac{0,05 \times 0,95}{50}} \approx 0,112 \text{ (11,2%) voor het percentage dat minder dan 8 uur werkt.}$$

Steekproefresultaat is 9/50 ofwel 18% en dat valt niet binnen de 95-procentgrenzen.

**Opgave 54**

$$\text{Rechtergrens 95%-gebied: } 0,65 + 2 \times \sqrt{\frac{0,65 \times 0,35}{150}} \approx 0,728 \text{ en } 0,728 * 150 = 109,2 \Rightarrow \text{er zullen dan}$$

minsten 110 mensen van de 150 moeten slagen.

**Opgave 55**

Bedenk: gemiddelde – s.d. is linkergrens van 68%-gebied van de levensduur en die grens is hier dus precies die 8000 branduren => 16 % van de lampen heeft dus (volgens de vuistregels van de normale verdeling) een levensduur die kleiner is dan 8000 uur waardoor ze niet voor de test zullen slagen.

Linkergrens 95%-gebied van het aantal dat de test niet doorstaat:

$$0,16 + 2 \times \sqrt{\frac{0,16 \times 0,84}{600}} \approx 0,13$$

Rechtergrens 95%-gebied van het aantal dat de test niet doorstaat:

$$0,16 + 2 \times \sqrt{\frac{0,16 \times 0,84}{600}} \approx 0,19$$

De steekproefproportie is 48/600 = 0,08 en dat valt NIET binnen het 95%-gebied.

