

1.5 Uitspraken

Inleiding

Een statistisch onderzoek is opgezet om uitspraken te kunnen doen. In kranten en tijdschriften staat het er bol van. Maar vaak ontbreekt belangrijke informatie: er staat bijvoorbeeld wel een gemiddelde, maar er wordt geen spreidingsmaat vermeld. Of er wordt niet vermeld hoe de steekproef is samengesteld...

Welke uitspraken kun je wel doen en welke niet? En wat moet je allemaal vermelden om de betrouwbaarheid van een uitspraak duidelijk te maken?

Een aantal opgaven uit dit onderdeel is afkomstig uit de NLT-module 'Maak het verschil', net als enkele datasets.

Je leert in dit onderwerp

- wat een klokvormige frequentieverdeling is en welke vuistregels daarbij horen;
- hoe je verantwoorde uitspraken kunt doen bij statistische gegevens;
- kritisch te kijken naar uitspraken en conclusies die je tegenkomt.

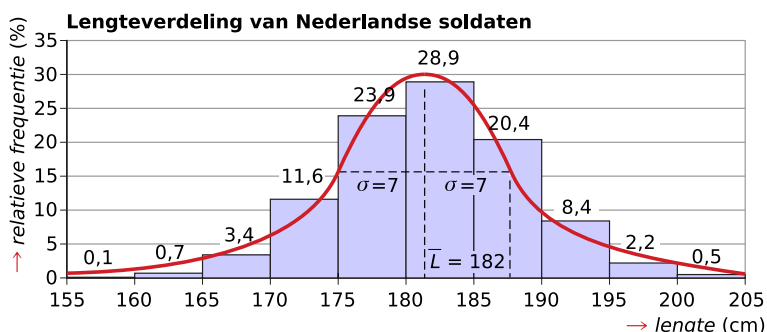
Voorkennis

- allerlei statistische diagrammen interpreteren en maken;
- centrummaten en spreidingsmaten berekenen.

Verkennen

Opgave V1

Bekijk de lengteverdeling van een grote groep mannelijke Nederlandse soldaten.

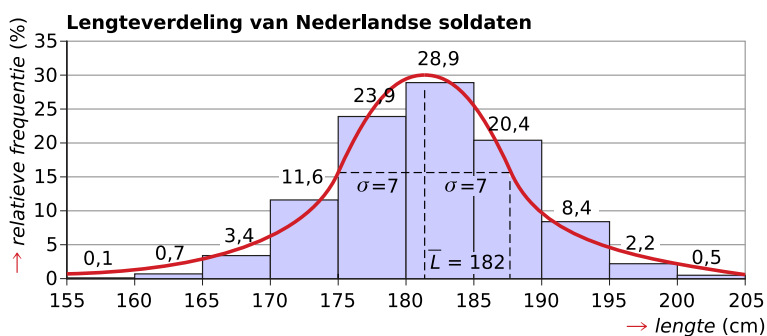


Figuur 1

Kun je verklaren waarom de lengtes van zo'n grote groep mannen de vorm van het silhouet van zo'n kerkklok heeft? Waar zit het gemiddelde?

Uitleg

Bekijk de lengteverdeling van een grote groep mannelijke Nederlandse soldaten.



Figuur 2

Het histogram of de bijbehorende frequentiepolygoon heeft een klokvorm. Bij veel continue variabelen, bijvoorbeeld bij gewicht, lengte of inhoud, krijg je zo'n klokvormige frequentieverdeling. Je kunt met behulp van gemiddelde en standaardafwijking twee algemene uitspraken doen. Deze uitspraken zijn vuistregels.

Vuistregel 1: tussen $\bar{x} - \sigma_x$ en $\bar{x} + \sigma_x$ zit 68% van de waarnemingsgetallen.

Vuistregel 2: tussen $\bar{x} - 2\sigma_x$ en $\bar{x} + 2\sigma_x$ zit 95% van de waarnemingsgetallen.

Deze uitspraken betreffen alleen de steekproef (beschrijvende statistiek). 68% van deze soldaten heeft een lengte tussen 175 cm en 189 cm. 95% van deze soldaten heeft een lengte tussen 168 cm en 196 cm.

Vaak wordt verondersteld dat deze gegevens voor de hele populatie Nederlandse jonge mannen gelden en dat die uitspraken op hen van toepassing zijn. Uitspraken doen over een populatie op grond van een steekproef kan alleen als die steekproef representatief is. En dat is hier nog maar de vraag.

Statistiek lijkt spijkerhard, maar je kunt sneller misleid worden door diagrammen en cijfers dan je denkt. Soms wordt een deel van een diagram of van een as weggelaten. Of de cijfers en uitspraken gaan over een te kleine of verkeerde steekproef. Wees altijd op je hoede met cijfers en diagrammen bij een onderzoek. Je hoort immers zelden dat uit een onderzoek geen conclusies getrokken kunnen of mogen worden.

Opgave 1

Gebruik de lengteverdeling van 90 meisjes die je vindt in het bestand [Lengteverdeling 90 meisjes](#).

- a** Maak van deze verdeling een frequentiepolygoon.

De frequentieverdeling is niet een perfecte klokvorm, omdat de steekproef veel te klein is.

- b** Bereken het gemiddelde en de standaardafwijking van de gegeven lengteverdeling.
c Onderzoek of voor de gegeven lengteverdeling de 68%-vuistregel geldt.
d Onderzoek of voor de gegeven lengteverdeling de 95%-vuistregel geldt.

Opgave 2

Op grond van een representatieve steekproef uit alle Nederlandse meisjes heeft een onderzoeksbureau geconcludeerd dat hun lengtes klokvormig verdeeld zijn met een gemiddelde lengte van 172 cm en een standaardafwijking van 6 cm.

- a** Bepaal met behulp van de vuistregels hoeveel procent van de Nederlandse meisjes langer is dan $172 + 6 = 178$ cm.
b Bepaal hoeveel procent korter is dan $172 - 2 \cdot 6 = 160$ cm.

Opgave 3

Je ziet hier een aantal conclusies uit de statistische gegevens. Geef commentaar op de uitspraken.

- a** In 1971 nam de NAVO 49% van alle militaire uitgaven voor haar rekening. In 1981 was dat nog 43%. De militaire uitgaven van de NAVO zijn in 1981 lager dan in 1971.
b Van alle verkeersongelukken op deze weg blijkt bij 25% alcohol een rol te hebben gespeeld. Rijden na het drinken van alcohol is veiliger dan rijden zonder alcohol.
c Wasmiddel XXX wast 20% witter dan alle andere wasmiddelen.
d School A heeft hogere percentages geslaagden dan school B. Je kunt beter op school A zitten als je snel wilt slagen.

Theorie en voorbeelden

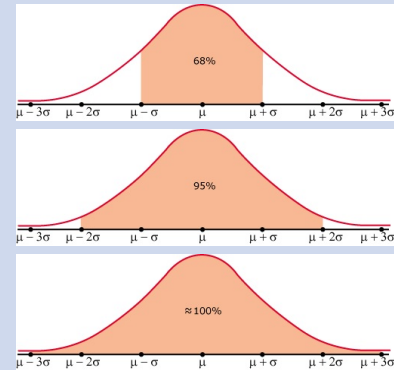
Om te onthouden

In elk deel van een boxplot zit 25% van de waarnemingen. Wanneer het histogram of de bijbehorende frequentiepolygoon bij benadering **klokvormig** is, zijn het gemiddelde \bar{x} en de standaardafwijking σ_x goede karakteristieken van de frequentieverdeling. Er gelden vuistregels.

- Vuistregel 1: tussen $\bar{x} - \sigma_x$ en $\bar{x} + \sigma_x$ zit 68% van de waarnemingsgetallen.
- Vuistregel 2: tussen $\bar{x} - 2\sigma_x$ en $\bar{x} + 2\sigma_x$ zit 95% van de waarnemingsgetallen.
- Vuistregel 3: tussen $\bar{x} - 3\sigma_x$ en $\bar{x} + 3\sigma_x$ zit bijna 100% van de waarnemingsgetallen.

Deze uitspraken betreffen de steekproef.

De uitspraken die je doet over je steekproef zijn alleen geldig voor de hele populatie als de steekproef een goede afspiegeling van die populatie is, dus **representatief** is. De uitspraken zijn betrouwbaarder als de steekproef voldoende groot is.



Figuur 3

Voorbeeld 1

De lengteverdeling van Nederlandse mannen boven de 20 jaar is bij benadering klokvormig. De gemiddelde lengte is 180,3 cm. De standaardafwijking is 7,74 cm.

Tussen welke twee lengtes zit volgens de vuistregels 68% van de Nederlandse mannen? En 95%?

Antwoord

Volgens de vuistregels zit 68% van deze mannen tussen de gemiddelde lengte min de standaardafwijking en de gemiddelde lengte plus de standaardafwijking. Dus 68% heeft een lengte tussen 172,6 cm en 188,0 cm.

De 95%-regel zegt dat 95% van de lengtes maximaal twee keer de standaardafwijking van het gemiddelde af zit. Dus 95% van de mannen heeft een lengte tussen 164,8 cm en 195,8 cm.

Opgave 4

Bekijk **Voorbeeld 1**. Neem aan dat voor de verdeling van lengte L van de Nederlandse mannen boven de 20 jaar deze verdeling geldt.

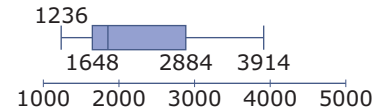
lengte Nederlandse mannen boven 20 jaar	
lengte	percentage
< 163	1
163– < 168	3,3
168– < 173	11,8
173– < 178	18,2
178– < 183	26,7
183– < 188	23,2
188– < 193	9,3
193– < 198	4,9
> 198	1,5

Tabel 1

Teken een bijpassend frequentiepolygoon en reken het gemiddelde \bar{L} en de standaardafwijking σ_L na. Laat zien dat er sprake is van een klokvorm.

Opgave 5

Voor een onderzoek naar de levensduur van een bepaald type batterijen is op basis van 200 waarnemingen een boxplot getekend.



Figuur 4

Geef aan welke van de volgende beweringen waar zijn. Licht je antwoord toe.

- A. Minimaal 25% van de batterijen gaat langer dan 3000 uur mee.
- B. Meer dan 50% van de batterijen heeft een levensduur van minder dan 2000 uur.
- C. De batterijen gaan gegarandeerd 1200 uur mee.
- D. Minstens 75% van de batterijen werkt nog na 1600 uur.

Opgave 6

Op de verpakking van een pak koffie staat een inhoud van 250 gram. In werkelijkheid is dat iets meer of minder. Het gewicht van 1000 pakken koffie wordt gemeten, zonder verpakking. Uit de metingen blijkt een gemiddeld gewicht van 254 gram. De standaardafwijking is 4 gram. Ga ervan uit dat de verdeling van het gewicht klokvormig is.

Geef aan welke uitspraken volgens de vuistregels waar zijn.

- A. Ongeveer 95% van de pakken koffie heeft een gewicht tussen 246 en 262 gram.
- B. Ongeveer 5% van de pakken koffie heeft een gewicht onder 246 gram.
- C. Ongeveer 16% van de pakken koffie heeft minder dan de beloofde 250 gram inhoud.
- D. Ongeveer 50% van de pakken koffie heeft een gewicht van 250 gram.
- E. Minimaal 75% van de pakken koffie heeft een gewicht van meer dan 250 gram.

Voorbeeld 2

Voorbeelden van misleidingen.

Belgen spreken langzamer dan Nederlanders

De schok was groot toen uit een artikel in *Onze Taal* bleek dat Belgen beduidend langzamer praten dan Nederlanders. In de Randstad haalt men 5,42 lettergrepen per seconde, in Oost-Vlaanderen slechts 4,43. Sommige mensen gingen het meteen controleren. De spreeknelheid van 21 miljoen Nederlandssprekenden werd bepaald door maar liefst 160 leraren en leraressen een stukje te laten opzeggen. Er waren acht groepen, dus twintig sprekers per groep. En dan werd ook nog gerapporteerd over het verschil tussen jong en oud, man en vrouw. Eén oude Antwerpse stotteraarster, en de achterstand is hopeloos. Onderzoeker Guy De Pauw maakte het allemaal nog erger door een dag later te verklaren dat de verschillen niet significant waren. Alsof dat er nog toe doet, met zo'n steekproef.
bron: vanmaanen.org, 2004, artikel Onze Taal, Hans van Maanen, wetenschapsjournalist

Vitalinea misleidt consument

In de nieuwe reclamespots voor het aanprijzen van Vitalinea van Danone gebruiken de reclameboys wel heel trieste, misleidende statistieken, waar de fouten zo van afdruipten. De reclame claimt dat 'Tijdens een studie bij 400 Belgen, is gebleken dat 80% van de deelnemers gemiddeld 3,6 kilogram afvalt.' Misschien valt je frank niet direct, maar deze statistiek wil helemaal niks zeggen. Waarom geven de onderzoekers het gemiddelde gewichtsverlies van slechts 80% van de deelnemers? Waarom niet van de volle 100%? Waar zijn de statistieken van die andere 20% deelnemers plots heen? Wat mij betreft zijn deze 20% mensen die niet meegeteld zijn allemaal 30 kilogram bijgekomen door het eten van Vitalinea, en komt het gemiddelde dus uit op een gewichtstoename bij het eten van Vitalinea. Het enige dat ik kan concluderen van de reclame, is dat als je wil vermageren, Vitalinea niet het goede product is. Simpele logica.
bron: anthony.lieken.net, 2005, Anthony Liekens

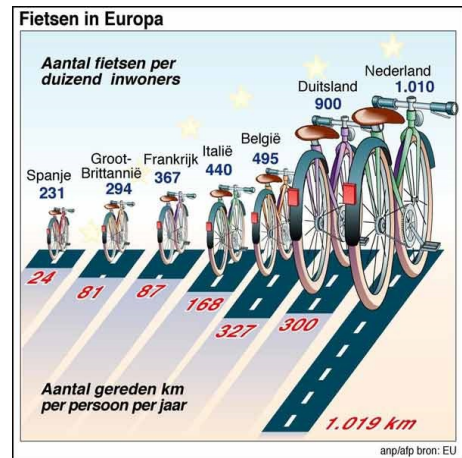
Opgave 7

In **Voorbeeld 2** zie je hoe slordige statistieken je kunnen misleiden en/of hoe soms slordige conclusies worden getrokken. Geef bij elk van de voorbeelden kort commentaar.

Opgave 8

Wij Nederlanders, we doen nauwelijks iets anders dan fietsen. Kijk maar.

- Welk land heeft vermoedelijk de meeste fietsen?
- Welke informatie ontbreekt om dit met zekerheid te kunnen zeggen?
- Hoeveel kilometer fietsen we per persoon in Nederland per jaar gemiddeld?
- Het rechter deel van de figuur is eigenlijk een staafdiagram. De lengte van de staaf past dan bij de grootte van de waarde die hij aangeeft. Klopt dat hier wel? Licht je antwoord toe met een voorbeeld.
- Waarom zijn, zo bekeken, de fietsen ook een staafdiagram? Klopt dat staafdiagram dan?



Figuur 5

Verwerken

Opgave 9

Bekijk het bestand **Gegevens 36 mannen**.

Op het moment van deze steekproef van 36 was de gemiddelde lengte van alle Nederlandse mannen 177,6 cm met een standaardafwijking van 6,6 cm. De verdeling van deze lengtes had een zuivere klokvorm.

gegevens van 36 mannen											
nr	leeftijd (j)	lengte (cm)	gewicht (kg)	nr	leeftijd (j)	lengte (cm)	gewicht (kg)	nr	leeftijd (j)	lengte (cm)	gewicht (kg)
i	L	l	G	i	L	l	G	i	L	l	G
1	20	180	95	13	46	180	95	25	67	171	68
2	23	184,5	91	14	47	177,5	75	26	67	170	81
3	23	180,5	90	15	51	196	86	27	68	169	97
4	29	169	91	16	52	188,5	93	28	68	176	72
5	30	176	80	17	53	170	85	29	71	171	84
6	33	182	82	18	54	178,5	77	30	71	180,5	75
7	36	182,5	85	19	56	172,5	74	31	71	172	77
8	36	176,5	74	20	61	164	85	32	74	171	77
9	38	180	73	21	61	181	89	33	75	183	95
10	40	190,5	112	22	61	173	95	34	75	172,5	74
11	41	184	89	23	61	181	89	35	82	173	66
12	44	179	75	24	63	181,5	100	36	84	175	90

Tabel 2

- Tussen welke twee waarden zou 68% van de lengtes van de 36 mannen moeten liggen als de steekproef representatief is?
- Ga na of deze steekproef voldoet aan de eerste vuistregel van de klokvormige frequentieverdeling.
- Tussen welke twee grenzen zou ongeveer 95% van de lengtes van de 36 mannen moeten liggen als de steekproef representatief is?
- Ga na of deze steekproef voldoet aan de tweede vuistregel van klokvormige frequentieverdeling. Bekijk daarvoor de tabel met gegevens van 36 mannen.

Opgave 10

Een kledingzaak maakt broeken die op maat afgeknipt worden. Zo zijn ze voor iedereen precies lang genoeg. Maar hoe langer de broek vóór het afknippen, hoe duurder en hoe meer stof er wordt weggegooid. De ontwerpers hebben laten onderzoeken wat de beenlengte van Nederlandse mannen is. Die is klokvormig verdeeld met gemiddelde 80 cm en standaardafwijking 5 cm. De langst gemeten lengte is 95 cm.

- a Alle broeken kunnen zo lang worden als de langste man nodig heeft. Hoe lang worden de broeken dan? Noem een bezwaar tegen dit idee.

Een andere mogelijkheid is om de lengte van de broek zo te kiezen dat 84% van de mannen erin past.

- b Hoe lang moet de broek dan worden?

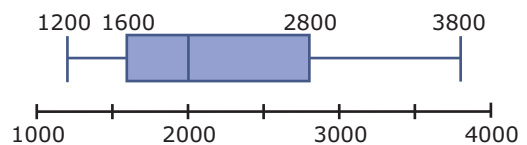
De damesafdeling van de kledingzaak wil ook zulke broeken. De verdeling van de beenlengtes van vrouwen is ook klokvormig, met $\bar{x} = 74$ cm en $\sigma_x = 4$ cm.

- c Hoe lang moeten de vrouwenbroeken zijn zodat 84% van de vrouwen erin past?

Opgave 11

In een bedrijf is het modale salaris ongeveer € 1600,00 per maand. Het gemiddelde salaris is € 1800,00 per maand. Het hoogste salaris is dat van de algemeen directeur. In de boxplot zie je de verdeling van de salarissen over alle 120 mensen die bij het bedrijf werken. Bereken in de volgende gevallen steeds het modale salaris en het gemiddelde salaris en teken de nieuwe boxplot. Doe voor elk van de drie situaties een kenmerkende uitspraak over de gevolgen van de maatregel voor de laagstbetaalde 25% werknemers.

- a Alle medewerkers krijgen een loonsverhoging van 3%.
- b Alle medewerkers krijgen een maandelijkse toeslag van € 200,00.
- c Het salaris van de algemeen directeur wordt met € 800,00 per maand verhoogd.



Figuur 6

Opgave 12

Bekijk de gegevens van pasgeboren kinderen in Nederland. De verdeling is klokvormig. Doe vier uitspraken met behulp van de vuistregels over geboortegewicht en geboortelengte.

Geboortelengte in cm						
< 45,5	45,5 -< 47,5	47,5 -< 49,5	49,5 -< 51,5	51,5 -< 53,5	53,5 -< 55,5	> 55,5
3,7	6,6	18,1	37,5	22,9	6,9	4,2
Geboortegewicht in kg						
< 1,5	1,5 -< 2,0	2,0 -< 2,5	2,5 -< 3,0	3,0 -< 3,5	3,5 -< 4,0	4,0 -< 4,5
0,8	1,0	3,4	12,4	32,2	33,6	12,6

Figuur 7

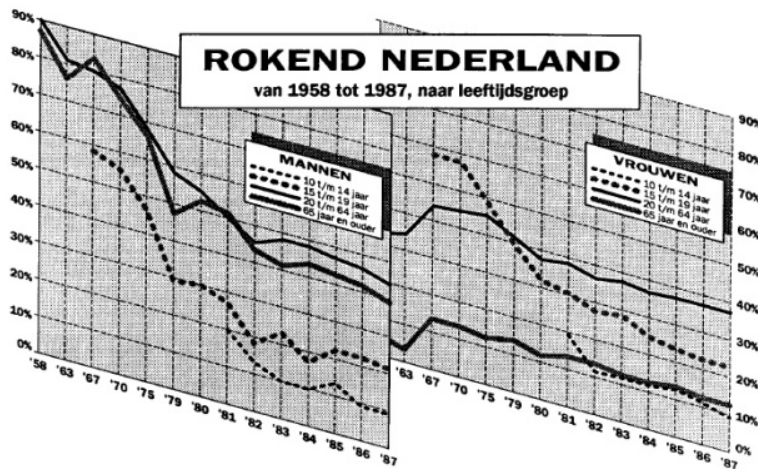
Opgave 13

Open het Excel bestand [Patiëntgegevens](#).

- a Bereken de gemiddelde lengte van zowel de vrouwelijke als de mannelijke patiënten en de bijbehorende standaardafwijkingen. Is er verschil tussen de lengtes van mannen en vrouwen?
- b Onderzoek of 50% van de mannen langer is dan de 84% kortste vrouwelijke patiënten.

Opgave 14

De lijndiagrammen komen uit een krantenartikel uit 1988. Volgens de linker grafiek rookte in 1958 nog 90% van de mannen in de leeftijdsgroep van 20 tot 65 jaar. In 1987 was dit gedaald tot 43%. Deze sterke daling wordt door de tekenaar op een misleidende wijze benadrukt.



Figuur 8

- Wat veroorzaakt deze misleiding?
- Bekijk het diagram van de mannen van 15 tot 20 jaar. De grafiek ziet er ook voor de jaren 1982 tot 1987 dalend uit. Daalt het percentage rokers van die categorie ook werkelijk?
- Bij welke van deze acht diagrammen is er vrijwel nooit van daling sprake?

In het krantenartikel stond:

Een overzicht van de rookgewoonten in Nederland in 1987 gaf, net als in de jaren daarvoor, opnieuw een daling te zien van het aantal rokers in ons land. Hoewel de betrekkelijk snelle daling in de jaren zeventig en het begin van de jaren tachtig is afgenomen, heeft die tendens zich de afgelopen drie jaar gestabiliseerd op een daling van 1% per jaar. Kon in 1958 worden becijferd dat 60% van de Nederlandse mannen en vrouwen in de leeftijdsgroep van 15 tot 65 jaar rookte, volgens cijfers van de Stichting Volksgezondheid en Roken was dat in 1987 afgenomen tot 37%.

Een lezer van dit artikel denkt dat die 37% niet kan kloppen. Hij redeneert zo:

- de laatste drie jaar was er een daling van 1%;
 - volgens de tekst en de figuur was de daling in de periode daarvoor nog sterker;
 - in 1958 was het percentage rokers 60%;
 - in de 29 jaar van de periode 1958-1987 is daar zeker $29 \cdot 1\% = 29\%$ van af gegaan, dus in 1987 moet het minder dan 31% zijn.
- Leg uit waarom 37% wel correct kan zijn als je de 1% daling per jaar goed interpreteert.

Toepassen

In 1947 zijn bij 5001 vrouwelijke klanten van de Bijenkorf vijftien lichaamsmaten opgemeten. Vervolgens is gekeken welke van deze maten het meest bruikbaar zijn om een **maatsysteem voor kleding** op te baseren. Bekijk een deel van de uitkomst van het onderzoek in het bestand [Statistiek Bijenkorf 1947](#).

Je kunt onder andere met behulp van de **vuistregels voor klokvormige verdelingen** uitspraken doen over deze populatie. En (er van uitgaande dat deze steekproef representatief was voor de bezoekers van de Bijenkorf in die tijd) een maatsysteem voor vrouwenkleding bedenken.

Opgave 15

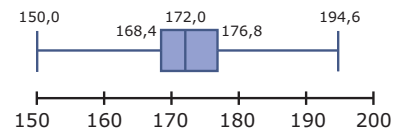
Bekijk de uitkomst van het onderzoek in het bestand [Statistiek Bijenkorf 1947](#).

- Bepaal de gemiddelde lengte en de bijbehorende standaardafwijking.
- Maak het bijpassende histogram voor de variabele *lengte*. Laat zien dat er van een mooie klokvormige frequentieverdeling sprake is.
- Hoeveel lengtes verschillen meer dan één keer de standaarddeviatie van het gemiddelde? Hoeveel procent van de vrouwen betreft dit?
- Hoeveel procent van de lengtes verschilt meer dan twee keer de standaardafwijking met het gemiddelde?
- Komen deze antwoorden overeen met de vuistregels voor klokvormige verdelingen?
- Waarom heeft het geen zin om na te gaan of voor de frequentieverdeling van de gewichten de vuistregels opgaan?

Testen

Opgave 16

Je ziet een boxplot van de lengtes van 1064 vaders van ongeveer 100 jaar geleden.



Figuur 9

- Welke uitspraak kun je doen over de 25% kortste mannen?
- Controleer de uitspraak: 266 van deze 1064 mannen had een lengte vanaf 172,0 tot 176,8 cm.
- Geef twee redenen waarom de lengtes van deze mannen geen zuiver klokvormige frequentieverdeling hebben.

Opgave 17

Je ziet de leeftijdsopbouw van leraren in het havo/vwo in procenten.

- Bereken voor elk van de vijf genoemde jaren het gemiddelde en de standaarddeviatie van de leeftijden van deze leraren.
- Teken de vijf frequentiepolygonen en geef die waarden daarin aan.
- Welke conclusies kun je trekken?
- De waarden van 1995 en 2000 zijn schattingen die de onderzoekers in 1994 hebben gedaan. Passen die schattingen bij de gegevens uit de voorgaande jaren?

Leraren in HAVO/VWO					
Leeftijd	1980	1985	1990	1995	2000
-29	20	12	8	3	6
30-34	21	19	13	8	4
35-39	18	21	19	14	9
40-44	14	18	21	20	14
45-49	11	13	19	22	20
50-54	8	10	14	18	22
55-59	5	6	5	12	17
60-64	3	1	1	3	8

Tabel 3

(bron: 'Onderwijsrecht Gelders onderzoek' - Arbon 1994)

Opgave 18

Open het bestand [Etmaaltemperaturen De Bilt](#).

- Maak een histogram van de temperaturen in de maand juli over de jaren 1755 tot 1900. Neem een klassenbreedte van 1 °C.
- Maak ook een histogram voor de periode van 1900 tot 2007.
- Vergelijk de twee histogrammen met elkaar. Kun je concluderen dat de temperatuur in de maand juli na 1900 gemiddeld hoger is dan in de voorgaande periode?



© 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

Met de Math4All Foliostaat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@math4all.nl een gratis inlog voor de maatwerkdienst aanvragen.
