

3.6 Totaalbeeld

Samenvatten

Je hebt nu het onderwerp **Statistiek** doorgewerkt. Er moet een totaalbeeld van deze leerstof ontstaan... Ga na, of je al de bij dit onderwerp horende begrippen kent en weet wat je er mee kunt doen. Ga ook na of je de activiteiten die staan genoemd kunt uitvoeren. Maak een eigen samenvatting!

Begrippenlijst

- kwalitatieve variabelen — discrete/continue kwantitatieve variabelen
- (relatieve) frequentie — frequentietabel, frequentieverdeling — somfrequenties — klassenindeling
- staafdiagram, lijndiagram, cirkeldiagram — histogram, (som)frequentiepolygoon — steelblad diagram
- centrummaten: modus, mediaan, gemiddelde — spreidingsmaten: spreidingsbreedte, kwartielafstand, standaardafwijking — kwartielen, boxplot
- klokvormige frequentieverdeling — twee vuistregels

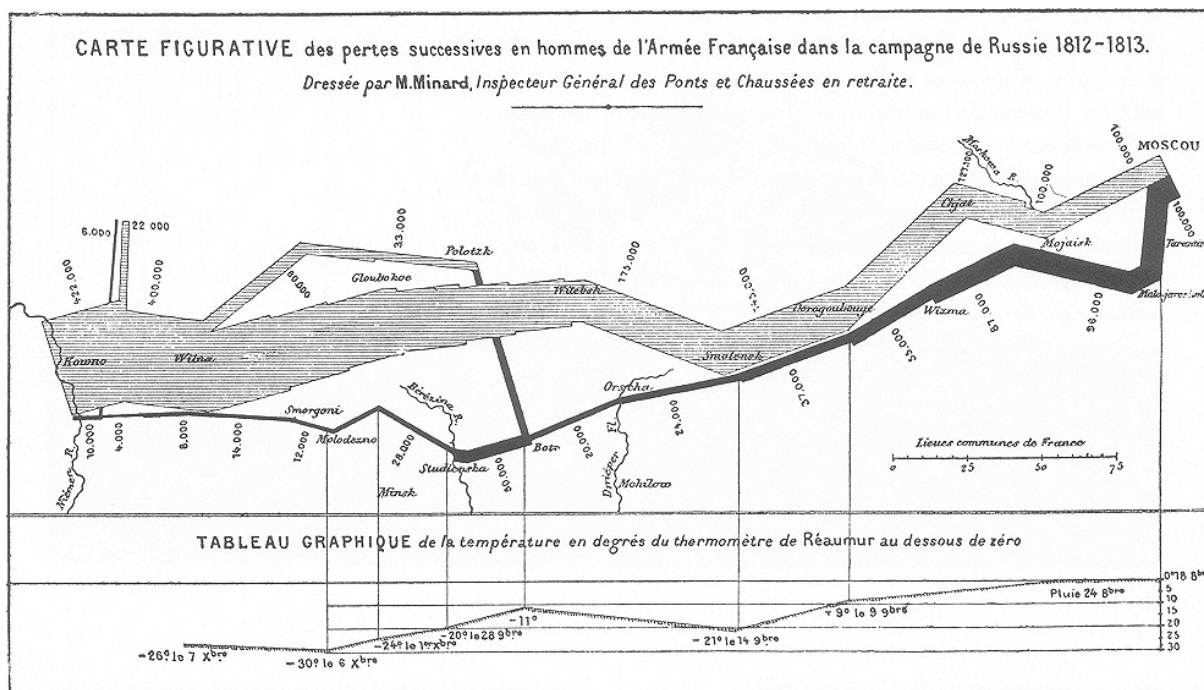
Activiteitenlijst

- representativiteit van een steekproef beoordelen — soorten statistische variabelen herkennen
- (som)frequentieverdelingen maken — werken met klassen
- diagrammen gebruiken om statistische resultaten weer te geven (ook met de GR en/of Excel)
- centrummaten en spreidingsmaten berekenen (schatten) zowel met als zonder klassenindeling (ook met de GR en/of Excel)
- uitspraken doen op grond van steekproefverdelingen

Achtergronden

De beschrijvende statistiek ontstond pas toen staten en centrale overheden een belangrijke rol gingen spelen. Immers: vooral centrale overheden zijn geïnteresseerd in globale overzichten van de bevolking, omdat het aantal mensen waarover zij verantwoordelijk is te groot is om van iedereen alles te weten. Via **Historische hoogtepunten van de grafische verwerking** kun je veel informatie vinden over het ontstaan van statistieken en de mensen die daarin een rol hebben gespeeld.

Hieronder zie je Minard's beroemde diagram van de tocht van Napoleon naar Rusland.



Figuur 1

Testen

Opgave 1

In 1987 verscheen het geruchtmakende boek 'Women and Love: A Cultural Revolution in Progress' van Shere Hite. De auteur beschreef daarin de resultaten van een onderzoek onder 100000 Noord-Amerikaanse vrouwen met betrekking tot hun relatie. Van de vrouwen die de vragenlijst terugstuurden:

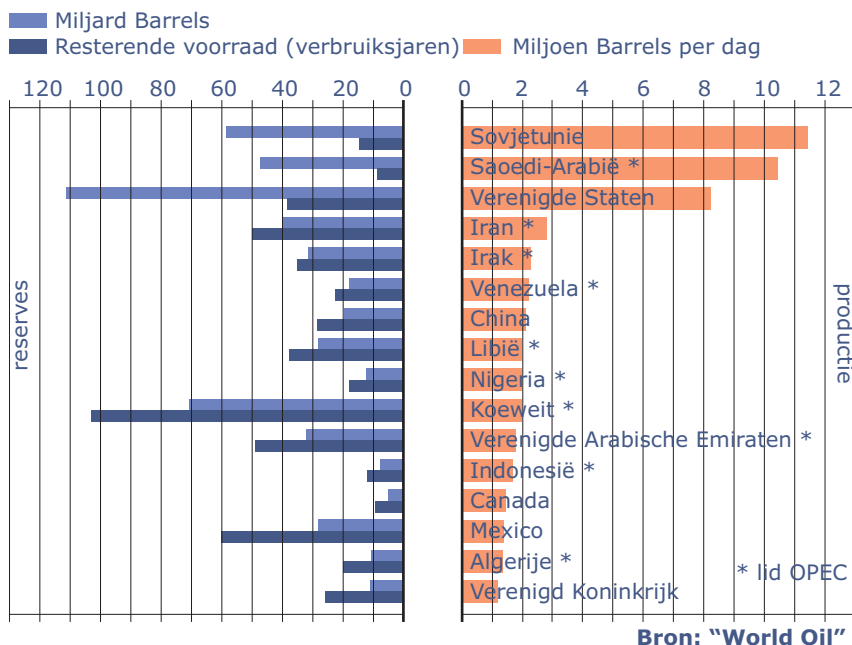
- voelde 84% zich emotioneel niet goed in hun relatie;
- had 95% psychische of lichamelijke mishandeling doorstaan;
- gaf slechts 13% aan na twee jaar huwelijk nog van hun man te houden.

Om te laten zien dat Hite een representatieve steekproef had opgesteld, gebruikte ze tabellen. Hite hield onder meer rekening met ras/afkomst en de regio. Zo'n 4500 vrouwen stuurden de vragenlijst ingevuld terug aan de auteur.

- Specificiteer zo nauwkeurig mogelijk welke populatie Shere Hite met behulp van deze steekproef wilde onderzoeken en geef aan met welk type statistiek Shere Hite bezig was. Met beschrijvende of met verklarende statistiek?
- Shere Hite heeft rekening gehouden met jaarlijks inkomen, type woonplaats, regio en met ras/afkomst van de onderzochte vrouwen. Op welke kenmerken had ze haar steekproef nog verder kunnen uitsplitsen om de steekproef nog representatiever te maken? Noem er twee.
- Uit de gegevens die je nu over dit onderzoek hebt, kun je niet afleiden of deze steekproef aselekt is. Definieer de term aselechte steekproef voor dit onderzoek.
- Kon Shere Hite met 100% zekerheid weten dat 87% van de vrouwen na twee jaar huwelijk niet meer van hun man hield? Licht je antwoord toe.

Opgave 2

In de figuur vind je gegevens over de oliereserves en de olieproductie uit 'Aardolie, de halve wereld draait erop', van de Stichting School en Bedrijf.



Figuur 2

- Voor welk type variabele zie je staafdiagrammen omtrent olie afgebeeld?
- Hoeveel procent van de reserves, gerekend in vaten, is in handen van de OPEC-landen?
- Wat is de productieproportie van Rusland?

- d Geef zowel de reserves (in miljarden vaten) als de productie (in miljoenen vaten per dag) weer in een cirkeldiagram. Neem als categorieën de OPEC-landen, de Verenigde Staten, Rusland en overig.
- e Vergelijk de oliereserves en de olieproductie met elkaar op basis van je cirkeldiagrammen.

Opgave 3

Op de verpakking van een literpak melk staat 'Inhoud 1 liter'. In werkelijkheid wil dat nog wel eens iets meer of minder zijn. Uit metingen blijkt een gemiddelde inhoud van 1,002 liter. De standaardafwijking is 0,004 liter. De verdeling van de inhoud is klokvormig. Geef met behulp van de vuistregels bij de uitspraken aan of ze waar of niet waar zijn.

- a Ongeveer 95% van de pakken melk heeft een inhoud van meer dan 1,010 liter.
- b Ongeveer 32% van de pakken melk bevat minder dan 0,998 of meer dan 1,006 liter.
- c Ongeveer 50% van de pakken melk bevat 1,002 liter.

Stel je voor dat er iets mis is met de vulmachine van de melkpakken en dat de frequentieverdeling van de inhoud opeens scheef naar rechts wordt in plaats van klokvormig.

- d Beargumenteer of de volgende uitspraak wel of niet met zekerheid kan worden gedaan: 'Méér dan 50% van de melkpakken heeft een inhoud die groter is dan het (nieuwe) gemiddelde.'

Opgave 4

De tabel geeft de leeftijden van het personeel van twee bedrijven.

- a De twee bedrijven willen fuseren. Werknemers vanaf 60 jaar kunnen gebruikmaken van een afvloeiingsregeling. Hoeveel procent van de werknemers van het gefuseerde bedrijf kan daar gebruik van maken?
- b Wat is er (theoretisch) mis met de frequentietabellen van bedrijf A en bedrijf B?
- c Leg uit waarom een cumulatieve relatieve frequentiepolygoon van dergelijke frequentietabellen de zuiverste indruk geeft. Waarom geven een (relatieve) frequentiepolygoon, een histogram en een cirkeldiagram hier een vertekend beeld?
- d Wil je deze twee frequentietabellen alleen gebruiken om ze met elkaar te vergelijken, dan doen de bezwaren er minder toe. Teken de relatieve histogrammen van beide bedrijven met gebruikmaking van de gegeven klassenindeling en benoem de verschillen en overeenkomsten tussen beide frequentieverdelingen die je uit deze histogrammen mag en kunt afleiden.

leeftijden van het personeel			
bedrijf A		bedrijf B	
leeftijd	aantal	leeftijd	aantal
15– < 18	89	15– < 18	67
18– < 25	133	18– < 25	205
25– < 40	414	25– < 40	801
40– < 55	249	40– < 55	562
55– < 60	68	55– < 60	265
60– < 65	47	60– < 65	100
	1000		2000

Tabel 1

Opgave 5

Uit onderzoek van het gemengde boerenbedrijf bleek het houden van kippen een belangrijke rol te spelen bij het tot stand komen van het inkomen van deze boeren. Daarom werd de boeren gevraagd naar het aantal kippen op hun bedrijf.

<i>aantal kippen</i>	<i>aantal bedrijven</i>	<i>aantal kippen</i>	<i>aantal bedrijven</i>
1 – 10	5	101 – 110	123
11 – 20	12	111 – 120	101
21 – 30	19	121 – 130	85
31 – 40	24	131 – 140	79
41 – 50	33	141 – 150	60
51 – 60	52	151 – 160	43
61 – 70	69	161 – 170	21
71 – 80	75	171 – 180	9
81 – 90	108	181 – 190	4
91 – 100	120	191 – 200	2

Tabel 2

- Met welk type variabele heb je hier te maken?
- Teken een cumulatief relatief frequentiepolygoon.
- Schat de mediaan en de beide kwartielen. Teken een boxplot bij deze gegevens.
- Je kunt het gemiddelde en de standaarddeviatie schatten vanuit de klassenmiddens en de frequentietabel. Laat zien hoe dat gaat en geef je antwoord in één decimaal nauwkeurig.
- Als het aantal kippen op een gemengd boerenbedrijf een klokvormige verdeling kent, hoeveel kippen hebben dan de 2,5% gemengde boerenbedrijven met de meeste kippen?

Toepassen

Opgave 6: Cooper-test

In de sport wordt veel met statistieken gewerkt. Er wordt namelijk nogal wat gemeten... De Cooper-test is een bekende manier om de conditie te meten. Je meet dan hoever je kunt hardlopen in 12 minuten. En de afgelegde afstand zegt iets over je fitheid.

Dit **XL-bestand met Coopertest resultaten** van een groep scholieren levert een verzameling basisgegevens. In deze tabel zie je hoeveel iemand onder de 30 jaar kan lopen met een conditie die slecht / redelijk / goed / zeer goed is.

Nu kun je nagaan hoeveel procent van deze groep scholieren in welke categorie valt, voor mannen en vrouwen afzonderlijk of voor beide groepen samen. Ook kun je de gemiddelde score, de standaardafwijking, e.d., berekenen. En nagaan of er van een klokvormige verdeling sprake is, de vuistregels kloppen... Maar het is natuurlijk leuker om dit met resultaten van de eigen school te doen.

	<i>vrouwen</i>	<i>mannen</i>
slecht	0 – < 1800	0 – < 2000
redelijk	1800 – < 2200	2000 – < 2400
goed	2200 – < 2700	2400 – < 2800
zeer goed	> 2700	> 2800

Tabel 3

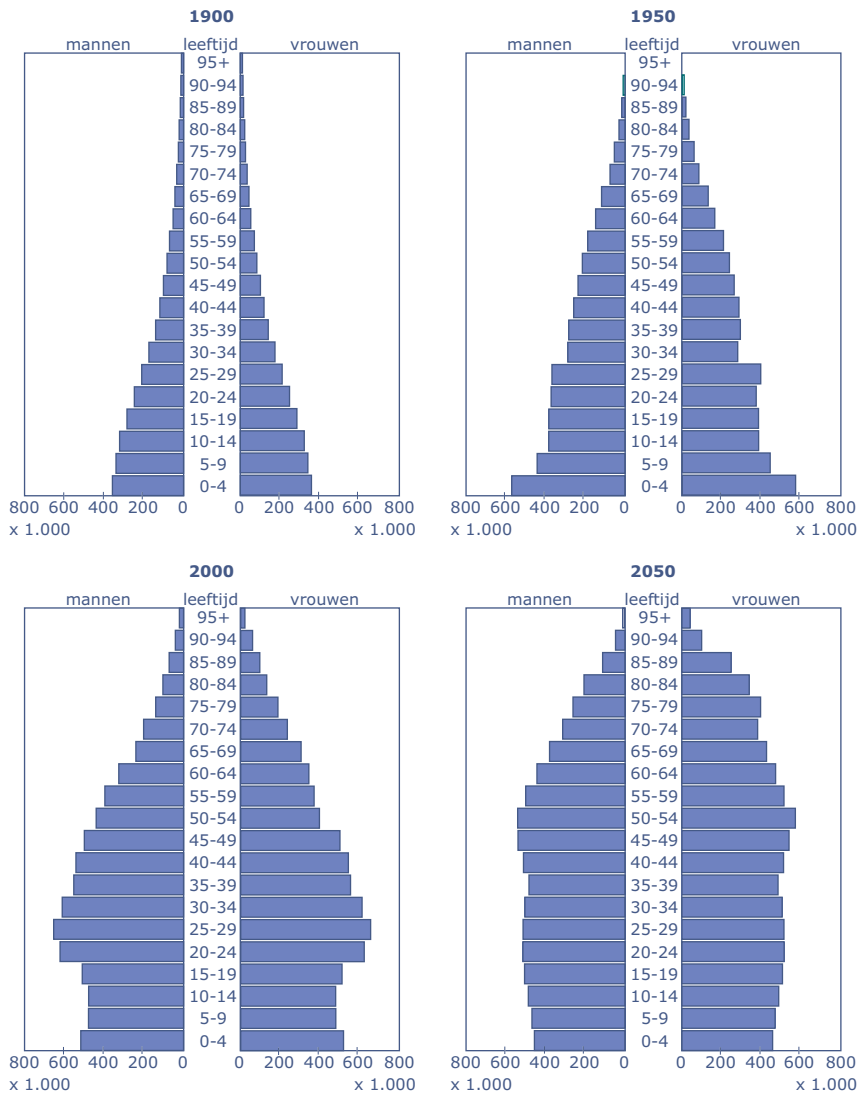
In 2004 volbrachten 182 leerlingen in klas 4 en 5 een Cooper-test. Van de leerlingen die de Coopertest (12 minutenloop) uitliepen is de afgelegde afstand genoteerd. Misschien kun je ook werken met gegevens van je eigen school!

- Welke klassenindeling zou je kiezen voor de vrouwen en welke voor de mannen om te kunnen concluderen of de conditie slecht, redelijk, goed of zeer goed is?
- Verwerk de gegevens voor vrouwen en mannen apart in twee afzonderlijke histogrammen.

- c** Verwerk de gegevens voor de vrouwen in een cirkeldiagram. Verwerk de gegevens voor de mannen in een cirkeldiagram.
- d** Welk percentage vrouwen heeft een goede of zeer goede conditie? Welk percentage mannen heeft een goede of zeer goede conditie?
- e** Typeer de afgelegde afstanden van de mannen en de vrouwen bij de Coopertest in gemiddelde, modus en mediaan.
- f** Teken van de resultaten van de mannen en van de vrouwen een boxplot. Bereken de kwartielafstand en de spreidingsbreedte.
- g** Doe een uitspraak op grond van beide boxplots waarin je de conditie van de mannen en de vrouwen met elkaar vergelijkt.
- h** Bereken voor beide groepen de standaardafwijking.
- i** Van hoeveel procent van de mannen en van hoeveel procent van de vrouwen wijkt de gelopen afstand meer dan de standaardafwijking van het gemiddelde af?
- j** Van hoeveel procent van de mannen en van hoeveel procent van de vrouwen wijkt de gelopen afstand meer dan twee keer de standaardafwijking van het gemiddelde af?
- k** Doe een uitspraak op grond van deze laatste gegevens bij waarin je de conditie van de mannen en de vrouwen met elkaar vergelijkt.

Opgave 7: Leeftijdsdiagrammen NL

Je ziet hier vier leeftijdsdiagrammen van Nederland.



Bron: Bevolkingsprognose Midden Variant

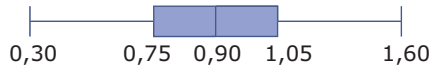
Figuur 3

- Wat voor soort diagrammen zijn dit?
- Hoeveel kinderen van 0–4 waren er in 1900 ongeveer? En in 1950?
- In de jaren rond 1950 zijn er nogal veel kinderen geboren. Waaraan kun je dat zien?
- Waaraan kun je zien dat Nederland aan het vergrijzen is?
- Waaraan kun je zien dat vrouwen gemiddeld ouder worden dan mannen?
- De belangrijkste centrummaat bij een leeftijdsdiagram is de modale klasse. Welke klasse was in 1900 de modale klasse? En in 2050?
- Welke gevolgen heeft het feit dat de modale klasse steeds hoger komt te liggen?
- De bevolkingsopbouw van Nederland is ook goed zichtbaar te maken in boxplots. Teken bij het leeftijdsdiagram van 1900 een boxplot (mannen en vrouwen samen). Doe dat ook bij het geschatte leeftijdsdiagram van 2050. Noem minstens drie karakteristieke verschillen en geef er verklaringen bij.

Examen

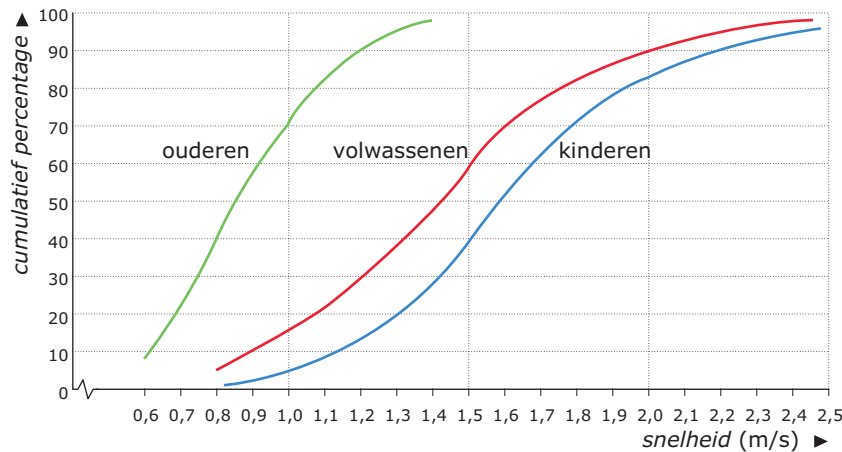
Opgave 8: Oversteken

Men heeft onderzoek gedaan naar de loopsnelheden van voetgangers. Bij dit onderzoek zijn de voetgangers in drie leeftijdsgroepen verdeeld, namelijk kinderen, volwassenen en ouderen. Met de gegevens uit het onderzoek heeft men een boxplot gemaakt voor de loopsnelheden van de groep ouderen.



Figuur 4

De snelheden die bij de boxplot vermeld zijn, zijn in meters per seconde. Meer gedetailleerde informatie over de groepen zie je in de volgende figuur.



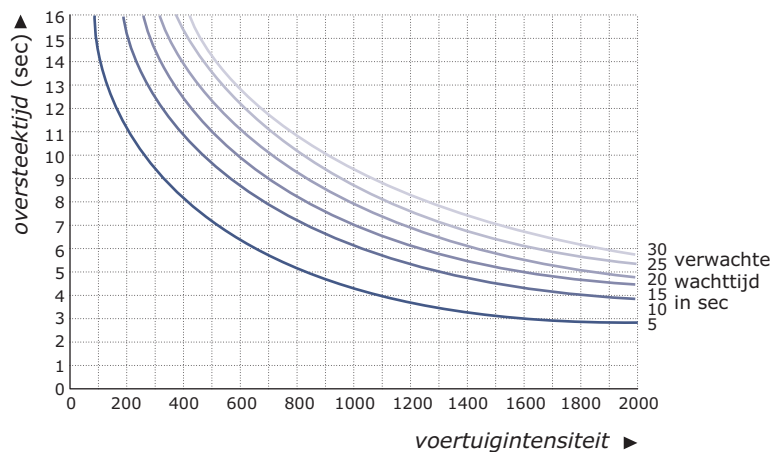
Figuur 5

Op de verticale as staat een cumulatief percentage; dit houdt in dat afgelezen kan worden hoeveel procent van de mensen van de verschillende groepen met de aangegeven snelheid of een lagere snelheid loopt. Zo kun je bijvoorbeeld aflezen dat voor de groep ouderen bij een snelheid van 1 m/s het cumulatieve percentage bijna 70 is. Dus bijna 70% van de ouderen loopt met een snelheid van 1 m/s of langzamer. Aan de hand van onder andere deze gegevens wordt een model gemaakt voor de tijd die de mensen nodig hebben om een weg over te steken. Neem aan dat de loopsnelheden ook voor het oversteken van een weg gelden. We bekijken het oversteken van een 20 meter brede weg. Er wordt recht overgestoken, dus men loopt daarbij 20 m.

- a** Maak met behulp van de gegevens uit het boxplot voor het oversteektijden van ouderen. Licht je werkwijze toe.

Tot nu toe hebben we alleen gekeken naar de tijd van oversteken zelf. Als je bij een weg aankomt, kun je niet altijd meteen oversteken; soms moet je een aantal seconden wachten. Deze wachttijd hangt samen met de drukte op de weg en de benodigde oversteektijd. De drukte op de weg wordt aangegeven met het aantal voertuigen dat per uur passeert (voertuigenintensiteit). Omdat ouderen in het algemeen minder snel lopen, zal voor deze groep de benodigde oversteektijd en dus ook de wachttijd groter zijn dan bijvoorbeeld voor kinderen.

Er is een model gemaakt voor de samenhang tussen oversteektijd, voertuigenintensiteit en verwachte wachttijd.



Figuur 6

In de figuur hierboven is dat voor zes verschillende wachttijden in beeld gebracht. Uit deze figuur is bijvoorbeeld af te lezen dat volgens dit model bij een oversteektijd van 9 s en een voertuigenintensiteit van 700 voertuigen per uur rekening gehouden moet worden met een wachttijd van 15 s.

- b** Teken de grafiek die het verband aangeeft tussen de oversteektijd en de verwachte wachttijd bij een voertuigenintensiteit van 800. Teken de grafiek alleen voor wachttijden van 5 tot en met 30 s.

We willen een beeld krijgen van de totale tijd die een rol speelt bij het oversteken van een weg van 20 m breed en een voertuigenintensiteit van 800 voertuigen per uur. We spreken dan over de somtijd. Als we iemands verwachte wachttijd en zijn oversteektijd optellen, krijgen we zijn somtijd. We bekijken nu de groep van volwassenen. De hoogste snelheid die in deze groep is waargenomen is 2,6 m/s.

- c** Wat is de langste somtijd en wat is de kortste somtijd van de 10% snelste volwassenen? Licht je antwoord toe.

Opgave 9: Vaders en zonen

De Engelsman Karl Pearson was een van de grondleggers van de moderne statistiek. Hij heeft zich vaak bezig gehouden met statistiek over biologische onderwerpen. Ongeveer een eeuw geleden onderzocht hij, samen met zijn collega Alice Lee, of in Engeland zonen gemiddeld langer zijn dan hun vaders. Zij vergeleken de lengtes van 1064 zonen en hun vaders. De zonen studeerden allen aan een Londonse universiteit.

- a** Is hier sprake van een aselechte steekproef? Licht je antwoord toe.

Hier zie je een overzicht van de resultaten. Elke stip stelt één vader-zoon-paar voor. De lengte van de vader staat op de horizontale as, de lengte van de zoon op de verticale as. De lengtes zijn gegeven in inches (1 inch = 2,54 cm). In de figuur is een lijn getekend. Als een stip op deze lijn ligt, dan zijn de vader en de zoon precies even lang. We noemen een vader en zijn zoon ongeveer even lang als ze minder dan 2 inch in lengte verschillen.

- b** Teken in de figuur op het **werkblad** het gebied waarin de punten liggen die horen bij vaders en zonen die ongeveer even lang zijn. Licht je werkwijze toe.

- c** Kun je met behulp van het getekende gebied concluderen dat de zonen gemiddeld langer zijn dan hun vaders? Licht je antwoord toe.

Op de bijlage zie je een boxplot van de lengtes van de 1064 vaders. De vijf kenmerkende getallen van de boxplot staan erbij. Op de bijlage vind je ook een lijst met de lengtes van alle 1064 zonen. De getallen in deze lijst staan op volgorde van grootte. Na iedere 10 getallen staat een streepje. Na iedere 50 getallen staat bij het streepje hoeveel getallen er tot daar staan.

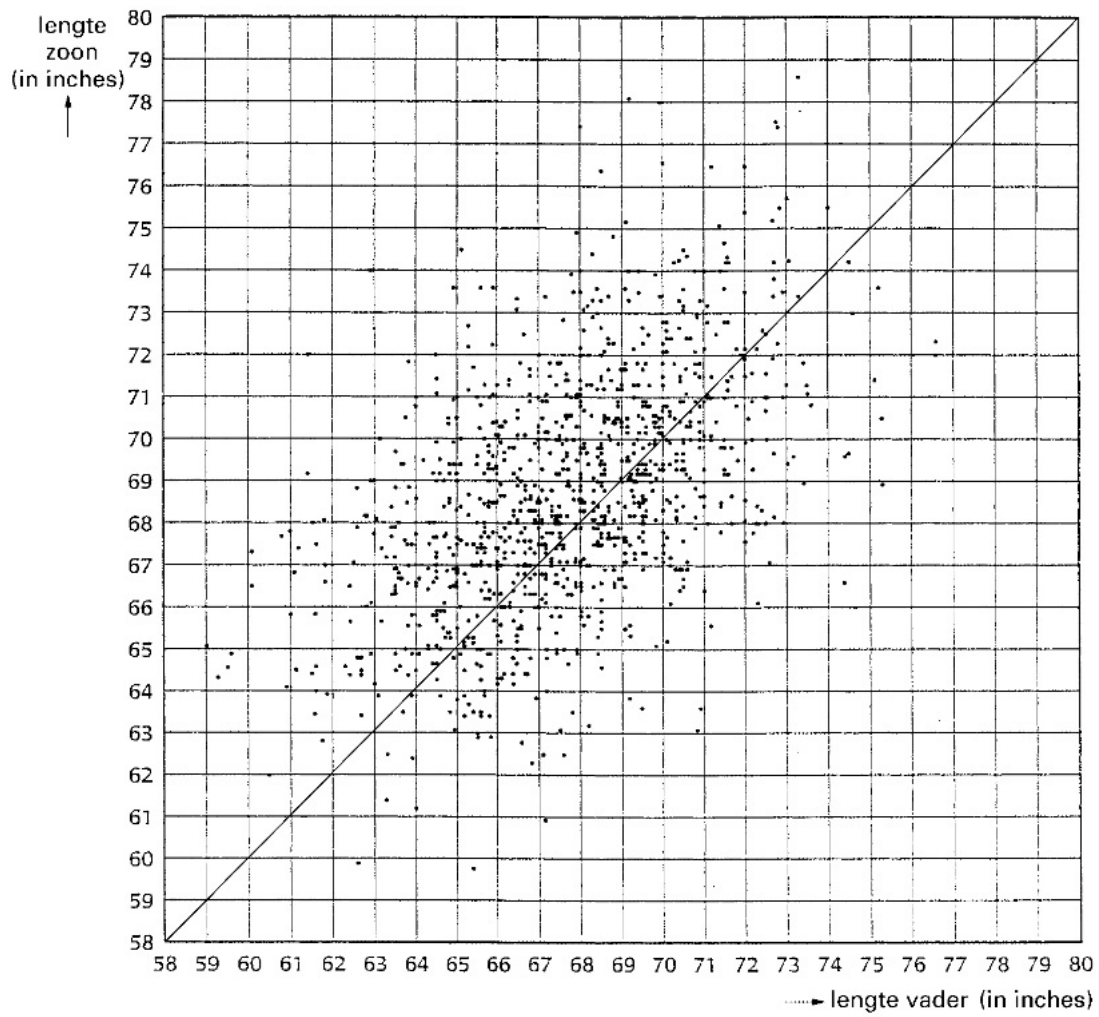
- d** Teken de boxplot van de lengtes van de zonen. Schrijf de vijf kenmerkende getallen van de boxplot erbij.

Het onderzoek dat bovenstaande getallen opleverde, is ongeveer honderd jaar geleden gedaan. In die tijd hadden jonge mannen een gemiddelde lengte van 68,6 inch. Dat is niet zo groot want 68,6 inch is maar 174 cm (1 inch = 2,54 cm). Tegenwoordig is de helft van de jonge mannen langer dan 182 cm. Honderd jaar geleden was veel minder dan de helft van de jonge mannen zo lang. De lengte was toen klokvormig verdeeld met een gemiddelde van 68,6 inch en een standaardafwijking van 2,7 inch.

- e Onderzoek met de vuistregels of het aantal jonge mannen langer dan 182 cm in die tijd groter of kleiner dan 16% van het totaal aantal jonge mannen was.

(bron: examen wiskunde A havo 2003, eerste tijdvak, aangepast)

Werkblad bij Opgave 9 op pagina 8





© 2022

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.

Email: f.spijkers@math4all.nl

Met de Math4All maatwerkdienst kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@xs4all.nl een gratis inlog voor de maatwerkdienst aanvragen.
