

4.5 Totaalbeeld

Samenvatten

Je hebt nu alle theorie van het onderwerp 'Conclusies trekken' doorgewerkt. Het is nu tijd om een overzicht over het geheel te krijgen.

Begrippenlijst

- kwalitatief, kwantitatief — discreet, continu — nominaal, ordinaal
- verschil van kwalitatieve variabelen — procentueel verschil — cumulatief verschilpercentage — kruistabel en phi
- verschil van kwantitatieve variabelen — boxplot en vuistregels — effectgrootte
- verband tussen variabelen — trendlijn

Activiteitenlijst

- vuistregels normale verdeling toepassen
- drie methoden om het verschil tussen kwalitatieve variabelen in een getal uit te drukken toepassen — vuistregels via formulekaart
- drie methoden om het verschil tussen kwantitatieve variabelen in een getal uit te drukken toepassen — vuistregels via formulekaart
- statistisch verband tussen twee variabelen herkennen — verschil statistisch verband en oorzakelijk verband herkennen

Achtergronden

Je hebt diverse manieren gezien om statistische variabelen te vergelijken of juist naar een statistisch verband tussen twee variabelen te zoeken. Wie hebben al die methoden toch bedacht?

- De *phi* voor het bepalen of in een twee bij twee kruistabel de twee kwalitatieve variabelen weinig, een beetje of veel van elkaar verschillen, is bedacht door **Karl Pearson (1857 - 1936)**, één van de grondleggers van veel statistische methoden. In feite hangt dit getal nauw samen met de correlatiecoëfficiënt r , in dit geval voor twee punten in een assenstelsel.
- De *effectgrootte* is een statistische maat voor hoe sterk het effect van een handeling is op een populatie, waarbij vergeleken wordt met een andere populatie waarop die handeling niet wordt toegepast ("controlegroep"). Deze maat is bedacht door de Amerikaanse statisticus en psycholoog **Jacob Cohen (1923 - 1998)**.
- De *boxplot* wordt gebruikt om twee kwantitatieve variabelen te vergelijken. Hij is voor het eerst geïntroduceerd door de Amerikaanse wiskundige **John Tukey (1915 - 2000)**. Tukey was één van de grondleggers van de data-analyse.
- De *correlatiecoëfficiënt* r voor het bepalen of er tussen twee kwantitatieve variabelen een (statistisch) verband bestaat, is ook bedacht door **Karl Pearson (1857 - 1936)**, hoewel de trendlijn zelf al ouder was en door **Carl Friedrich Gauss (1777 - 1855)** met de kleinste-kwadraten-methode is afgeleid.



Figuur 1 Karl Pearson

Testen

Opgave 1

Op de Universiteit van Leiden is een onderzoek gedaan onder een groep studenten. Hierbij werden onder andere de lichaamslengte en de armlengte van de studenten gemeten. Ook werden het geslacht, de kleur van de ogen en de voorkeurshand genoteerd. In totaal ging het om 34 mannelijke studenten en 32 vrouwelijke studenten. Enkele resultaten zie je in de tabel.

geslacht	voorkeurshand	
	links	rechts
man	5	29
vrouw	2	30

Tabel 1

Op basis van resultaten van dit onderzoek wil men uitspraken doen over alle mannelijke en vrouwelijke studenten in Leiden.

- Noem de variabelen uit de tabel en geef van elk van de variabelen aan of deze nominaal of ordinaal is. Licht je antwoord toe.
- Bepaal met behulp van de **Formulekaart** of het verschil in voorkeurshand tussen de mannelijke en de vrouwelijke studenten in het onderzoek groot, middelmatig of gering is.

In de tabel hieronder staan het gemiddelde en de standaardafwijking van de lichaamslengte van de studenten.

geslacht	lichaamslengte (cm)		
	aantal	gemiddelde	standaardafwijking
man	34	183,8	5,8
vrouw	32	170,8	8,1

Tabel 2

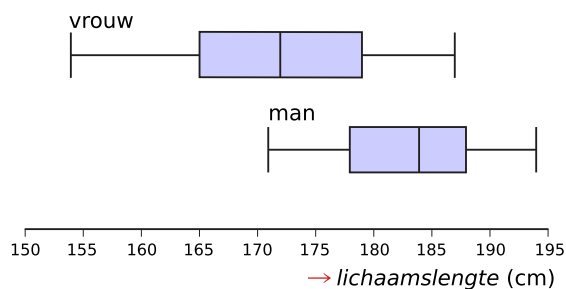
Op basis van deze tabel kunnen voor de lichaamslengte van mannen en vrouwen 95%-betrouwbaarheidsintervallen worden opgesteld.

- Onderzoek welk van deze twee intervallen het smalst is. Welke informatie geeft zo'n 95%-betrouwbaarheidsinterval in deze context?

Iemand die iets wil zeggen over de grootte van het verschil in lichaamslengte tussen de mannen en de vrouwen in het onderzoek, zou gebruik kunnen maken van de effectgrootte. Bij de berekening van de effectgrootte zijn de gemiddeldes en standaardafwijkingen van beide groepen van belang. Ga ervan uit dat het verschil van de gemiddeldes gelijk blijft.

- Leg uit op welke wijze grotere standaardafwijkingen de grootte van het verschil (groot, middelmatig, gering) kunnen beïnvloeden.

De resultaten van de lichaamslengtes werden ook in boxplots verwerkt. Zie figuur 1.



Figuur 1

Figuur 2

Je kunt met behulp van de gegevens over de lichaamslengtes in de tabel bij c onderzoeken hoe groot het verschil in lichaamslengte in het onderzoek is. Dit kun je ook doen met behulp van de gegevens in figuur 1.

e Toon met behulp van de **Formulekaart** aan dat je met deze twee gegevensbronnen een verschillende conclusie trekt over de grootte van het verschil in lichaamslengte tussen de mannelijke en de vrouwelijke studenten.

f Beargumenteer welk van beide conclusies het best te verdedigen is.

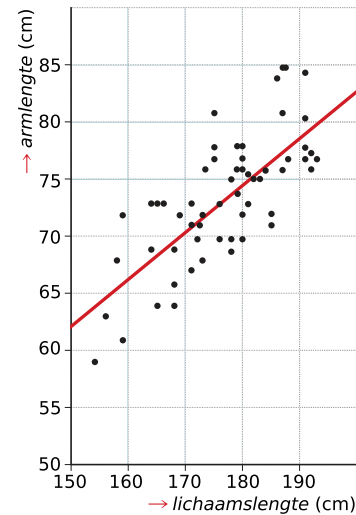
Om te onderzoeken of er een verband bestaat tussen de lichaamslengte en de armlengte van de studenten zijn de resultaten van het onderzoek in een spreidingsdiagram gezet. Zie figuur 2. Er blijkt in dit onderzoek bij benadering een lineair verband te zijn tussen de lichaamslengte L in centimeter en de armlengte A in centimeter. De getekende trendlijn geeft dit verband weer.

g Stel een formule op van de trendlijn.

De punten in het spreidingsdiagram liggen niet precies op de trendlijn. Ga ervan uit dat de armlengte normaal verdeeld is bij alle mogelijke lichaamslengtes tussen 150 cm en 200 cm en dat bij elke lichaamslengte de standaardafwijking van de bijbehorende armlengte gelijk is aan 3 cm. Als we de resultaten van dit onderzoek gebruiken om op basis van de lichaamslengte de armlengte van studenten te voorspellen, kunnen we hierbij figuur 2 gebruiken.

h Hoe kun je in figuur 2 het gebied waar de middelste 95% van de armlengten ligt voor elk van de lichaamslengten tussen 150 en 200 cm aangeven?

(bron: voorbeeldopgaven syllabus havo wiskunde A)



Figuur 2

Figuur 3

Opgave 2

In de Verenigde Staten is het gebruikelijk dat je in een restaurant een flinke fooi geeft aan degene die je bedient. Het basisloon is er zeer laag en daardoor is het bedienend personeel veel meer afhankelijk van fooien dan in Nederland. In New York bestudeerde een onderzoeker welke fooien er gegeven werden bij 500 rekeningen in restaurant A en 500 rekeningen in restaurant B.

In de figuur zijn de cumulatieve relatieve frequentiepolygonen van de fooien van de restaurants A en B getekend.

a In welk restaurant werd in totaal meer fooi gegeven? Licht je antwoord toe.

b In welk restaurant werden er relatief meer fooien tussen de 6 en de 8 dollar gegeven? Licht je antwoord toe.

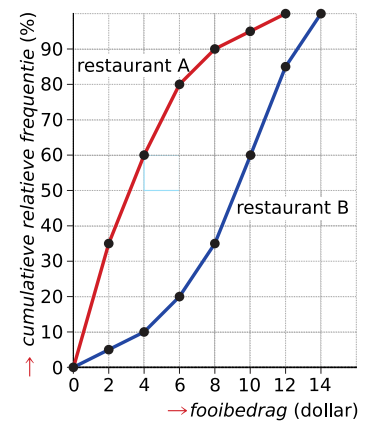
Ruim driekwart van de fooien in restaurant B is hoger dan de 75% laagste fooien van restaurant A.

c Leg uit hoe je dit in de figuur terug kunt vinden.

Op grond van de figuur kun je met behulp van de **Formulekaart** met twee verschillende vuistregels een uitspraak doen over het verschil tussen de fooien in restaurant A en die in restaurant B.

d Beschrijf hoe je met twee verschillende vuistregels dit verschil kunt bepalen en geef bij beide de conclusie: is het verschil groot, middelmatig of gering?

(bron: voorbeeldopgaven syllabus havo wiskunde A)



Figuur 4

Opgave 3

Vulkanen kunnen heel lang niet actief zijn. Dan zijn er geen erupties (uitbarstingen). Tijdens een actieve periode van een vulkaan zijn er wel erupties. Bij een eruptie komt er gesmolten steen, gas en as uit de vulkaan. Dat duurt een tijdje. Daarna is de vulkaan weer rustig, totdat de volgende eruptie begint.

Je bekijkt in deze opgave één actieve periode van één vulkaan. De actieve periode start bij de eerste eruptie. Wetenschappers hebben tijdens deze actieve periode gemeten:

- hoelang iedere eruptie duurt: de eruptieduur;
- hoelang de vulkaan rustig is tot de volgende eruptie begint: de tussentijd tot de eerstvolgende eruptie.

Tijdens deze actieve periode was de langste tijd tussen twee erupties 108 minuten. Na de allerlaatste eruptie was de vulkaan weer lange tijd rustig.

De eruptieduur is gemeten in tienden van een minuut nauwkeurig en de tussentijd in gehele minuten. De metingen zijn verwerkt in een puntenwolk. De eruptieduur E staat langs de horizontale as, de tussentijd tot de eerstvolgende eruptie T staat langs de verticale as.

Het meest linkse punt in de figuur hoort bijvoorbeeld bij een eruptie die 0,8 minuten duurde en waarna de vulkaan 49 minuten rustig was. In deze actieve periode zijn er 184 erupties geweest. De eerste 183 zijn weergegeven in het figuur. De allerlaatste eruptie van deze actieve periode duurde 1,7 minuten.

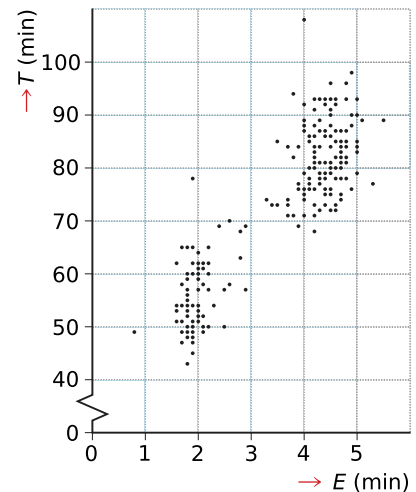
- a** De allerlaatste eruptie van een actieve periode kan niet in de puntenwolk worden weergegeven. Waarom niet?

In de figuur zijn de gegevens verwerkt van alle erupties, waarbij zowel de eruptieduur als de tussentijden tussen de erupties verwerkt zijn. Op basis van deze figuur is ook een schets te maken van de verdeling van de eruptieduur.

- b** Maak deze schets, met op de horizontale as de eruptieduur en op de verticale as het aantal erupties. Je hoeft hierbij niet het aantal erupties per eruptieduur exact te tellen, het gaat om een schets.

Iemand beweert op basis van deze schets van de eruptieduur te kunnen bepalen wat de standaardafwijking is van de eruptieduur. Hij doet dit door gebruik te maken van een van de vuistregels van de normale verdeling. Hij berekent het tijdsverschil tussen de langste eruptieduur en de kortste eruptieduur en deelt deze uitkomst door 6. De uitkomst van deze berekening beschouwt hij als een schatting van de standaardafwijking.

- c** Ben jij het met deze werkwijze eens? Geef duidelijk aan wat daarbij je argumenten zijn.
- d** Je kunt voor de 184 erupties de gemiddelde duur van de tussentijd T schatten. Met behulp van dit gemiddelde kun je een schatting maken van de lengte van de actieve periode van de vulkaan. Maak een schatting van de gemiddelde tussentijd tot de volgende eruptie en toon hiermee aan dat deze actieve periode van de vulkaan langer dan een week heeft geduurd.



Figuur 1

Figuur 5

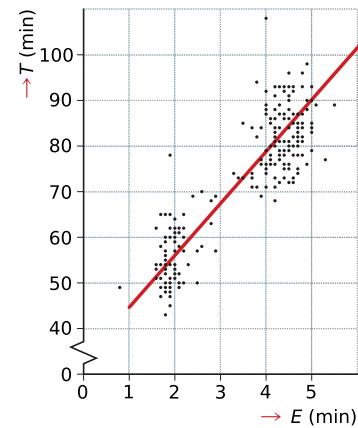
In figuur 2 is een lijn getrokken die zo goed mogelijk bij de metingen past. Deze lijn gaat door de punten (2,56) en (5,90). Met behulp van deze lijn kun je bij een gegeven eruptieduur een grove schatting maken voor de tussentijd die je daarbij kunt verwachten.

- e Bereken met behulp van de getrokken lijn uit dit figuur wat de geschatte tussentijd is bij een eruptieduur van 6 minuten.

Ga ervan uit dat de tussentijden bij een eruptietijd van 4,5 minuten normaal verdeeld zijn met een standaardafwijking van 4 minuten. Je kunt dan met een betrouwbaarheid van 95% aangeven tussen welke twee waarden de tussentijd tot de eerstvolgende eruptie zal liggen bij een eruptieduur van 4,5 minuten.

- f Geef deze twee waarden.
- g Piet beweert op basis van de gegevens uit de figuur te kunnen concluderen dat de eruptieduur groter wordt naarmate de tussentijd tussen twee erupties groter is. Harm is het niet met Piet eens. Geef een argument dat Harm kan gebruiken om Piet te overtuigen van zijn ongelijk.

(bron: voorbeeldopgaven syllabus havo wiskunde A)

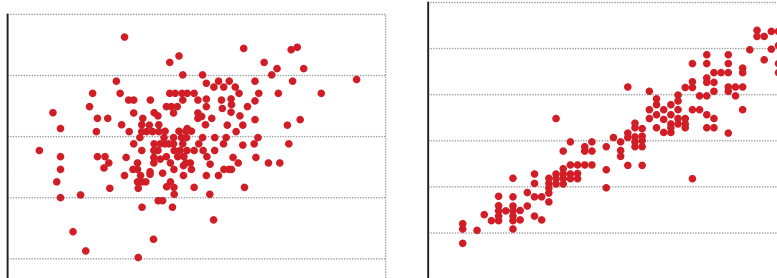


Figuur 2

Figuur 6

Opgave 4

In een onderzoek onder 199 echtparen is onder meer gevraagd naar de lengte en de leeftijd van de man en de vrouw. Er werd onder andere onderzocht of er bij bepaalde eigenschappen van de gehuwden sprake was van een bepaalde statistische samenhang. Dit heeft geresulteerd in de volgende twee puntenwolken:



Figuur 7

De rechter puntenwolk heeft betrekking op de leeftijden van de twee huwelijkspartners, waarbij de gegevens van de man op de horizontale as zijn uitgezet en die van de vrouw op de verticale as. De linker puntenwolk heeft betrekking op de lengte van beide partners. Ook hier zijn de gegevens van de man weer op de horizontale as uitgezet.

- a Waarom kun je door de rechter puntenwolk het beste een trendlijn tekenen?
- b Op het [werkblad](#) is de rechter figuur getekend met schaalverdelingen op beide assen. Teken de trendlijn in de figuur.

In de tabel is een aantal kentallen weergegeven uit het onderzoek.

	leeftijd man (jaar)	leeftijd vrouw (jaar)	lengte man (cm)	lengte vrouw (cm)
gemiddelde	42,6	40,7	173	160
minimum	20	18	156	141
maximum	64	64	195	176
standaardafwijking	11,6	11,4	6,9	6,2

Tabel 3

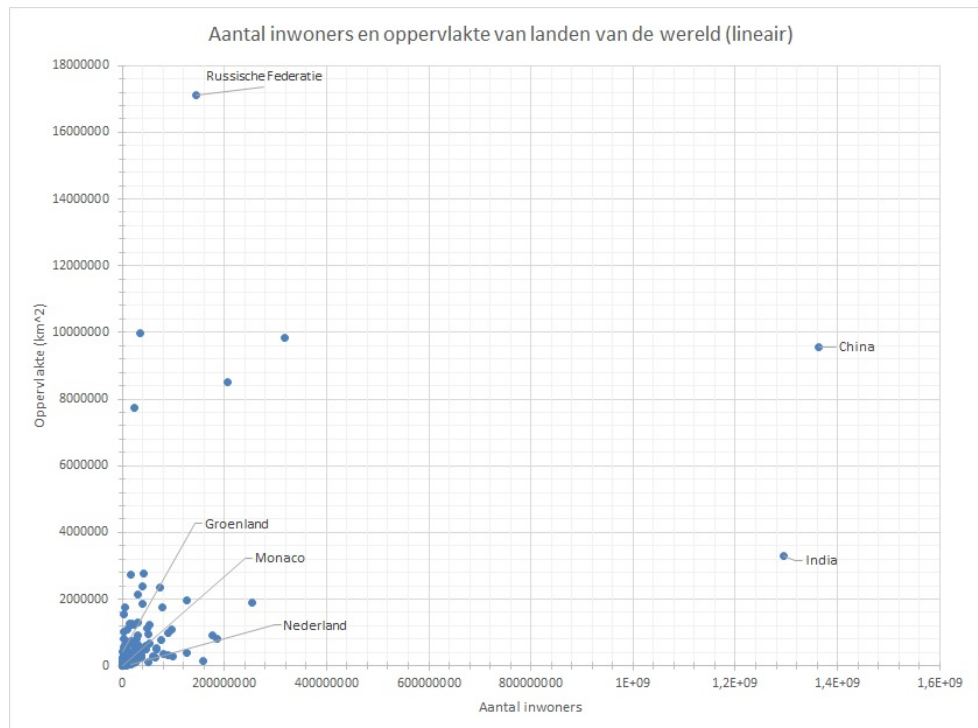
- c Bereken voor de leeftijden en de lengtes de effectgrootte om het verschil tussen mannen en vrouwen te vergelijken.

Gebruik de **Formulekaart** en rond af op twee decimalen.

(naar: voorbeeldopgaven syllabus havo wiskunde A)

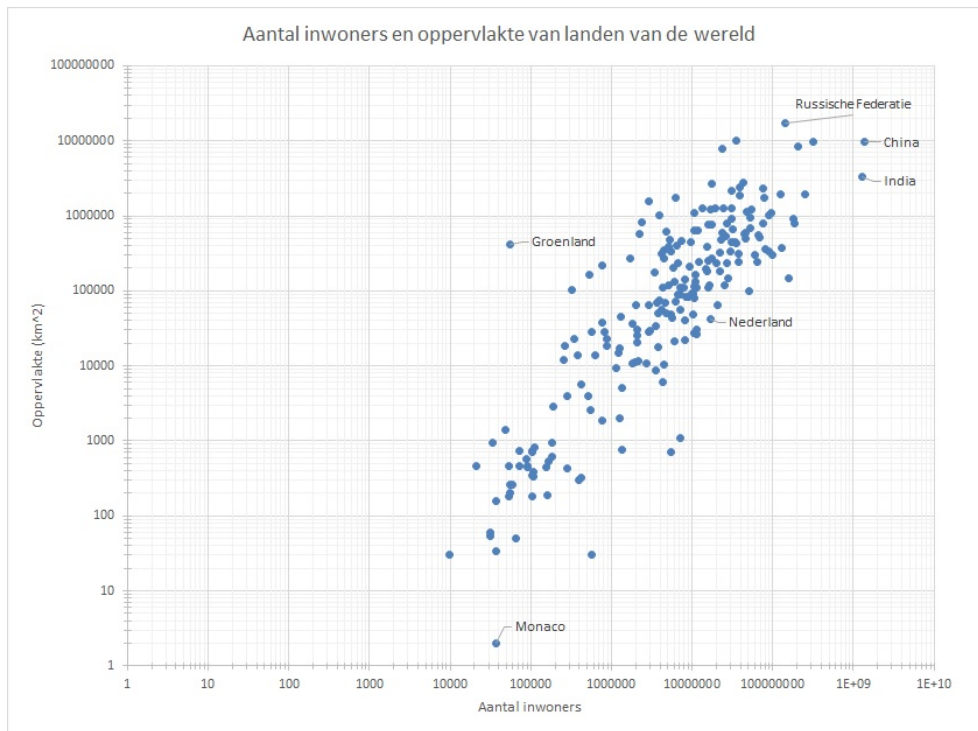
Toepassen

Opgave 5: Trendlijn op logaritmisches papier



Figuur 8

In de figuur hierboven voor een groot aantal landen het aantal inwoners en de oppervlakte weergegeven. Het is duidelijk dat aflezen hier erg moeilijk is. Een manier om dit te verbeteren is het gebruiken van logaritmische schalen. Dat zie je in de volgende figuur.



Figuur 9

- a** Het land dat de meeste inwoners telt, is China.
Lees af hoeveel inwoners China heeft. Geef je antwoord in miljarden, afgerond op één decimaal.
- b** Lees af hoe groot het oppervlak van China is. Geef je antwoord in miljoenen, afgerond op één decimaal.
- c** Het land met de grootste oppervlakte is Rusland.
Bereken het aantal inwoners per km² voor China en Rusland.
In de figuur is ook de trendlijn getekend.

De vergelijking van de trendlijn is: $A = \frac{1}{10000} \cdot P^{\frac{4}{3}}$

waarin A de oppervlakte in km² en P de bevolkingsomvang is.

- d** Bereken met deze formule het aantal inwoners van een land van 200000 km².
Geef je antwoord in miljoenen, afgerond op één decimaal.

Examen

Opgave 6: Alutech

Bij het bedrijf Alutech, dat aluminium buizen produceert, werkt men aan één stuk door van maandagmorgen tot en met zaterdagmiddag. De werknemers van dit bedrijf werken in ploegdiensten van 8 uur. Daarbij zijn er twee dagdiensten, genaamd A en B, en één nachtdienst N. Zie tabel.

werktijden Alutech							
	ma	di	wo	do	vr	za	zo
07.00 - 15.00 uur	A	A	A	A	A	A	-
15.00 - 23.00 uur	B	B	B	B	B	-	-
23.00 - 07.00 uur	N	N	N	N	N	-	-

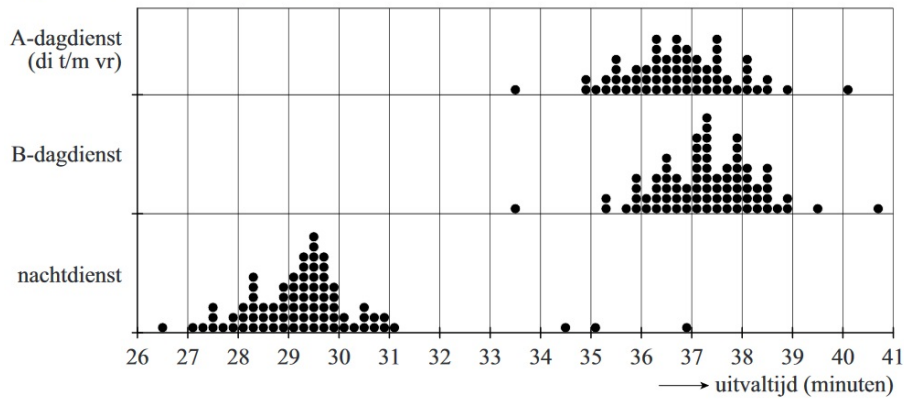
Tabel 4

Op maandagmorgen is men de eerste 3 uur van de A-dienst bezig om het productieproces op te starten en op zaterdagmiddag wordt het laatste uur van de A-dienst gebruikt om het proces gecontroleerd stil te zetten. Alle overige tijd wordt beschouwd als mogelijke productietijd.

- a Bereken het aantal uren mogelijke productietijd per week.

Tijdens elke dienst komen er storingen voor. Het productieproces wordt dan een aantal minuten stilgelegd totdat de storing verholpen is. Telkens wordt bijgehouden hoe lang de storing duurt. Na afloop van de dienst wordt de totale tijd van alle storingen genoteerd. Deze tijd noemt men de uitvaltijd. De directie wil dat de uitvaltijd zo klein mogelijk is. Om te onderzoeken hoe groot de uitvaltijd is, heeft men van 16 werkweken de uitvaltijden van de dag- en nachtdiensten vergeleken. Daarbij heeft men bij de A-diensten alleen gekeken naar de diensten die op dinsdag tot en met vrijdag vallen. De resultaten staan in de dotplot in figuur 1.

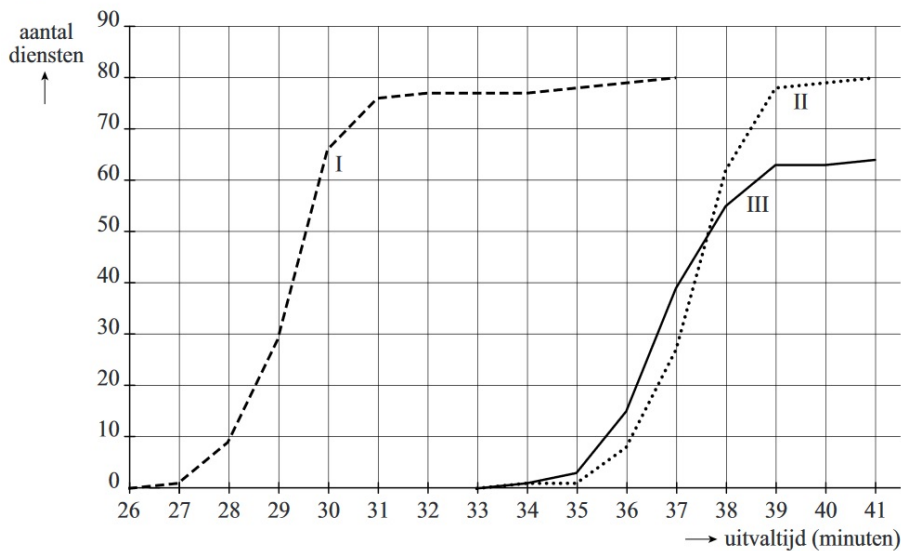
figuur 1 uitvaltijd per dienst



Figuur 10

Met de gegevens uit de dotplot zijn van alle diensten cumulatieve frequentiepolygoenen gemaakt. Er is gebruikgemaakt van een klassenindeling met klassenbreedte 1 minuut. Zie figuur 2.

figuur 2



Figuur 11

- b Beredeneer met behulp van figuur 1 welke cumulatieve frequentiepolygoon hoort bij dagdienst B.

Tot verbazing van de directie is de uitvaltijd tijdens de nachtdiensten in vrijwel alle gevallen lager dan tijdens de beide dagdiensten. Omdat elke minuut uitvaltijd het bedrijf geld kost, besluit de directie nader onderzoek te doen naar de uitvaltijden van de drie verschillende diensten. De directie berekent de mediaan, het eerste kwartiel en het derde kwartiel van de uitvaltijd voor elk van de drie diensten.

Zie deze tabel.

uitvaltijd per dag- of nachtdienst in minuten			
	mediaan	eerste kwartiel	derde kwartiel
dagdienst A (di-vr)	36,7	36,1	37,5
dagdienst B	37,3	36,6	37,9
dagdienst N	29,3	28,5	29,7

Tabel 5

Eerst kijkt de directie alleen naar het verschil tussen dagdienst A en dagdienst B. Met behulp van boxplots kun je een uitspraak doen over het verschil tussen de uitvaltijden van de twee dagdiensten. Daarvoor hoeven de boxplots niet getekend te worden.

- c Bepaal met behulp van de **Formulekaart** en de tabel of het verschil in uitvaltijd tussen dagdienst A en dagdienst B groot, middelmatig of gering is.

In figuur 1 is te zien dat er bij alle diensten waarnemingen zijn die opvallend afwijken van de rest. We leggen daarom vast wat we in deze opgave met een uitschieter bedoelen: “Een uitschieter is een waarneming die meer dan 1,5 keer de interkwartielafstand onder het eerste of boven het derde kwartiel ligt.”

- d Bepaal hoeveel waarnemingen bij dagdienst B uitschieters zijn.

Van elk van de drie verschillende diensten worden ook de gemiddelde uitvaltijd en de standaardafwijking berekend. Zie deze tabel.

uitvaltijd per dag- of nachtdienst in minuten		
	gemiddelde	standaardafwijking
dagdienst A (di-vr)	36,75	1,10
dagdienst B	37,29	1,04
dagdienst N	29,39	1,53

Tabel 6

Men vermoedt dat de lagere uitvaltijden tijdens de nachtdiensten te maken hebben met het feit dat de energietoevoer gedurende de nacht constanter is dan overdag. Daarom wordt de energietoevoer overdag verbeterd. De directie van Alutech hoopt daarmee de uitvaltijden van de dagdiensten terug te dringen.

Na verloop van tijd blijkt dat de gemiddelde uitvaltijd van de A-diensten van dinsdag tot en met vrijdag gelijk geworden is aan de gemiddelde uitvaltijd van de nachtdiensten. Ook de gemiddelde uitvaltijd van alle B-diensten is gelijk geworden aan de gemiddelde uitvaltijd van de nachtdiensten. De standaardafwijkingen van de A-diensten en B-diensten zijn niet veranderd.

Je kunt nu voor dagdienst B de oude met de nieuwe situatie vergelijken.

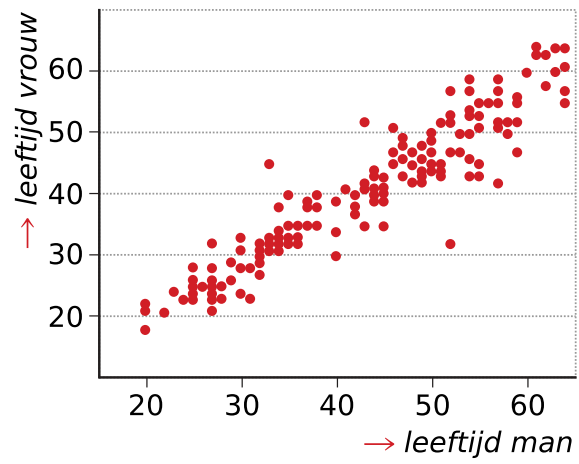
- e Bereken met behulp van de **Formulekaart** voor dagdienst B of het verschil in uitvaltijd tussen de oude en de nieuwe situatie groot, middelmatig of gering is.

Doordat de gemiddelde uitvaltijd van de A-diensten van dinsdag tot en met vrijdag en van alle B-diensten is teruggedrongen tot het gemiddelde van de uitvaltijd van de nachtdiensten, zal de productietijd toenemen.

- f Bereken met hoeveel uur de totale netto productietijd op jaarbasis toeneemt. Neem hierbij aan dat men 51 volledige werkweken per jaar werkt. Laat hierbij de A-diensten van maandag en zaterdag buiten beschouwing.

(bron: examen havo wiskunde A in 2017, tweede tijdvak)

Werkblad bij Opgave 4 op pagina 5.





© 2021

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.
