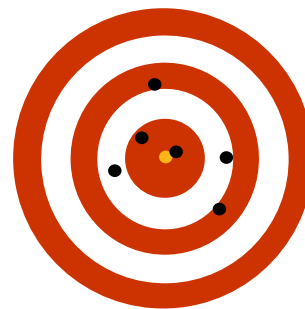


1.3 Centrum en spreiding

Inleiding

Om een verzameling gegevens, een dataset, zo simpel mogelijk weer te geven, gebruik je getallen die het centrum en de mate van spreiding van een frequentieverdeling weergeven. Je kent ze al: modus, mediaan en gemiddelde zijn centrummaten. Spreidingsbreedte of variatiebreedte en kwartielafstand zijn bekende spreidingsmaten.



Figuur 1

Je leert in dit onderwerp

- opnieuw data samenvatten met behulp van modus, mediaan en gemiddelde als centrummaten en met behulp van de variatiebreedte en de kwartielafstand als spreidingsmaten;
- opnieuw boxplots gebruiken om dezelfde soort gegevens van verschillende groepen te vergelijken (ook met MS-Excel).

Voorkennis

- werken met statistische gegevens, (frequentie)tabellen en diagrammen;
- werken met MS-Excel voor het maken van tabellen en diagrammen.

Verkennen

Opgave V1

Thijmen en Sven vergelijken de resultaten die ze het afgelopen jaar voor wiskunde hebben gehaald:

Sven	6,4	6,3	6,5	6,4	6,4	6,4	6,5	6,5	6,3
Thijmen	8,9	3,5	6,3	9,3	3,5	5,8	7,2	4,7	8,5

Tabel 1

- Bereken het gemiddelde van deze resultaten voor zowel Sven als Thijmen.
- Wie van de beide jongens heeft de grootste spreiding in zijn resultaten dit jaar? Geef uitleg.
- Wie van de beide jongens heeft volgens jou beter gepresteerd dit jaar? Geef uitleg.

Uitleg

Thijmen en Sven vergelijken de resultaten die ze het afgelopen jaar voor wiskunde hebben gehaald:

Sven	6,4	6,3	6,5	6,4	6,4	6,4	6,5	6,5	6,3
Thijmen	8,9	3,5	6,3	9,3	3,5	5,8	7,2	4,7	8,5

Tabel 2

Hun gemiddelden zijn hetzelfde. Kennelijk is het gemiddelde is niet voldoende om een goed beeld te krijgen van hun resultaten. Het gemiddelde is een centrummaat en zegt iets over het midden (het centrum) van een groep gegevens. Er zijn nog twee centrummaten, je kent ze al:

- de modus is de waarneming die het vaakst voorkomt;
- de mediaan is de middelste waarneming als je ze allemaal op volgorde zet.

Misschien zou je aan deze centrummaten iets kunnen zien, maar centrummaten alleen zeggen vaak niet zo veel. Je hebt een getal nodig dat iets zegt over de spreiding in de cijfers. Immers het grootste verschil is dat Sven veel constanter is en dat de cijfers van Thijmen een grote spreiding laten zien. Er zijn twee spreidingsmaten: de spreidingsbreedte en de kwartielafstand. Ook die ken je al, bekijk anders de **Theorie**.

Opgave 1

Bekijk in de **Uitleg** de gegevens van Sven en Thijmen nog eens.

- Bepaal van elk van hen het modale cijfer.
- Bepaal van elk van hen de mediaan van de cijfers.
- Waarom kun je nu nog steeds weinig zeggen als je hun resultaten wilt vergelijken?

Opgave 2

Je gaat nu de spreiding van de gegevens van Sven en Thijmen in de **Uitleg** bekijken.

- Bepaal van elk van hen de spreidingsbreedte.
- Bepaal van elk van hen de kwartielafstand van de cijfers.
- Waarom kun je nu wel wat meer zeggen als je hun resultaten vergelijkt?

Opgave 3

Je kunt de resultaten van Sven en Thijmen in de **Uitleg** ook indelen in klassen.

- Maak een klassenindeling waarvan de klassenmiddens de gehele cijfers 3, 4, ..., 9 zijn.
- Schat met behulp van deze klassenindeling het gemiddelde cijfer. Waarom wijkt dit af van het werkelijke gemiddelde?
- Bepaal hun modale gehele cijfers. Wat valt je op?
- Bepaal van beide frequentieverdelingen de spreidingsbreedte.

Theorie en voorbeelden

Om te onthouden

Je ziet hier een frequentietabel van de rapportcijfers voor wiskunde in klas 3G. Bij een **frequentieverdeling** van een aantal waarden (hier de cijfers) gebruik je de volgende **centrummaten**:

- het **gemiddelde**, dat is de som van alle waarden (rekening houdend met de frequenties) gedeeld door het totaal aantal waarden;
- de **modus**, dat is de waarde met de hoogste frequentie;
- de **mediaan**, dat is de middelste waarde.

Bij een **klassenindeling** gebruik je in plaats van de afzonderlijke waarden de **klassenmiddens**. Je kunt dan het gemiddelde en de mediaan alleen nog maar schatten. In plaats van de modus spreek je dan van de 'modale klasse'.

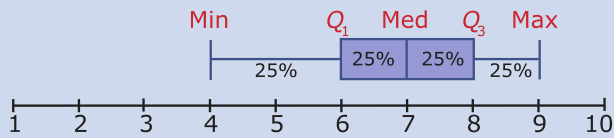
wiskunde	
RE	freq
1	0
2	0
3	0
4	1
5	4
6	9
7	8
8	6
9	3
10	0

Figuur 2

Omdat een centrummaat alleen weinig zegt, gebruik je daarnaast de volgende **spreidingsmaten**.

- de **variatiebreedte** of **spreidingsbreedte** is de grootste waarde min de kleinste waarde;
- de **kwartielafstand** is het derde kwartiel min het eerste kwartiel, waarin:
 - het **eerste kwartiel** de mediaan van de eerste helft van de waarden is,
 - het **derde kwartiel** de mediaan van de tweede helft van de waarden is.

De kleinste waarde en de grootste waarde, de mediaan en de twee kwartielen kun je overzichtelijk weergeven in een **boxplot**. Bij deze frequentietabel is de mediaan 7, het eerste kwartiel $Q_1 = 6$ en het derde kwartiel $Q_3 = 8$.



Figuur 3

Voorbeeld 1

In **dit Excel-bestand** vind je de rapportcijfers van klas 3G, onder andere die voor wiskunde.

Maak zelf een frequentietabel van deze cijfers zoals je die hiernaast ziet.

Bepaal het modale cijfer, het gemiddelde rapportcijfer zowel van de waarden op één decimaal nauwkeurig als van de afgeronde cijfers. Maak een boxplot van de cijfers voor wiskunde.

wiskunde	
RE	freq
1	0
2	0
3	0
4	1
5	4
6	9
7	8
8	6
9	3
10	0

Figuur 4

Antwoord

Bij **Toepassen** kun je nog even oefenen met het werken met Excel. Onder andere het bepalen van centrummaten en spreidingsmaten en het tekenen van boxplots.

Bij het berekenen van centrummaten en het maken van het boxplot moet je je goed bewust zijn van de waarden waar je mee rekent. Als je werkt met de afgeronde cijfers - dus eigenlijk met een klassenindeling en klassenmiddens - dan krijg je alleen schattingen van de werkelijke centrummaten en spreidingsmaten. In de praktijk is dat vaak wel goed genoeg. Bijvoorbeeld is het gemiddelde in beide gevallen 6,7 als je op één decimaal afrondt.

Opgave 4

Werk met de gegevens in **Voorbeeld 1**.

- Maak een frequentietabel van de gehele rapportcijfers. Waarom heeft het geen zin om een frequentietabel van de rapportcijfers op één decimaal nauwkeurig te maken?
- Bereken het gemiddelde van de gehele rapportcijfers en dat van de rapportcijfers op twee decimalen nauwkeurig.
- Waarom is de modus van de gehele cijfers een zinvol getal en de modus van de cijfers op één decimaal niet? Wat is het modale cijfer?
- Maak een boxplot van de gehele cijfers. Vergelijk dit met het boxplot van de cijfers op één decimaal dat Excel voor je maakt.

Opgave 5

Bekijk de tabel in **Voorbeeld 1**.

- Waarom is er bij de rapportcijfers sprake van een klassenindeling? Over welke klassen gaat het dan? De werkelijke cijfers zijn afgerond op één decimaal nauwkeurig. Ook dat zijn eigenlijk klassenmiddens.
- Van welke klassen?

Voorbeeld 2

Op de [website van het Centraal Bureau voor de Statistiek](#) (CBS) staan kun je gegevens vinden over de inkomens in Nederland. Voor 2011 staan daar de volgende gegevens op:

Inkomen per jaar (in euro)	Aantal personen (x 1000)
5000– < 10000	2490
10000– < 20000	2769
20000– < 30000	1963
30000– < 40000	1578
40000– < 50000	1304
50000– < 200000	2614

Tabel 3

Maak een schatting van het gemiddelde inkomen, bepaal het modale inkomen en maak een schatting van de mediaan.

Antwoord

Om het gemiddelde inkomen te schatten, maak je eerst een tabel van de klassenmiddens. Vervolgens vermenigvuldig je die met de frequenties (het aantal personen).

Klassenmiddens (in euro)	Aantal personen (× 1000)	Totaal per klasse
7500	2490	18675000
15000	2769	41535000
25000	1963	49075000
35000	1578	55230000
45000	1304	58680000
125000	2614	326750000
Totaal	12718	549945000

Tabel 4

Dus het gemiddelde inkomen is $549945000/12718 = 43241,47$ euro.

De modale klasse is de klasse die het meest voorkomt, dat is die met een inkomen van 10000– < 20000. Het midden van deze klasse is het modale inkomen.

Er zijn in het totaal 12718 (× 1000) gegevens, de middelste is 6359, deze bevindt zich in de klasse 20000– < 30000. Neem aan dat alle waarden gelijkmatig over de klassen zijn verdeeld. In de eerste twee klassen zitten al $2490 + 2769 = 5259$ (× 1000) waarden. Je moet dus de 1100e waarde van die klasse schatten:

$$20000 + 10000 \times \frac{1100}{1963} \approx 25604 \text{ euro.}$$

Opgave 6

Bekijk de inkomens tabel in [Voorbeeld 2](#).

- Waarom ontbreekt in de tabel een rij voor mensen die meer dan € 200.000 per jaar verdienen?
- Waarom is het twijfelachtig of de inkomens gelijkmatig over de klassen zijn verdeeld?
- Bereken ook de kwartielfstand uitgaande van een gelijkmatige verdeling en teken een bijpassend boxplot.
- Kun je op grond van het boxplot uitleggen waarom dit een scheve verdeling wordt genoemd?

Opgave 7

De vogelbescherming vraagt ieder jaar mensen met een tuin om te tellen hoeveel volgens er van verschillende soorten in hun tuin zijn. In 2010 was de top 10 als volgt samengesteld:

1	huismus	160119
2	koolmees	105838
3	merel	100705
4	vink	76790
5	pimpelmees	63547
6	spreeuw	61426
7	kauw	45924
8	Turkse tortel	44043
9	houtduif	37870
10	roodborst	34658

Tabel 5

- Hoe vaak komt een vogel uit de top 10 gemiddeld in de tuinen voor?
- Waarom heeft het geen zin om een mediaan of een modus te bepalen?

Verwerken

Opgave 8

Het ministerie van OC en W waar onderwijs onder valt houdt jaarlijks bij hoeveel vroegtijdige schoolverlaters er zijn. Een vroegtijdige schoolverlater stopt met zijn/haar opleiding voordat het diploma is behaald. Voor het schooljaar 2011 - 2012 waren dat de volgende aantallen.

leeftijd	≤ 13	14	15	16	17	18	19	20	21	22
aantal	253	385	503	1663	3645	10883	7783	5239	3533	2358

Tabel 6

Neem voor iedereen in de categorie ≤ 13 aan, dat de leeftijd 13 jaar is.

- Hoeveel bedraagt de gemiddelde leeftijd van een vroegtijdige schoolverlater?
- Hoeveel is de modale leeftijd van een vroegtijdige schoolverlater?
- Bepaal de mediaan.
- Wat vind je van deze resultaten?

Opgave 9

Het ministerie van OC en W waar onderwijs ondervalt houdt jaarlijks bij hoeveel vroegtijdige schoolverlaters er zijn. Een vroegtijdige schoolverlater stopt met zijn/haar opleiding voordat het diploma is behaald. Hier zie je de resultaten van twee schooljaren.

Neem voor iedereen in de categorie ≤ 13 aan, dat de leeftijd 13 jaar is.

- Bereken de spreidingsbreedte en de kwartielafstand bij de gegevens over vroegtijdige schoolverlaters over de schooljaren 2005-2006 en 2011-2012.
- Trek een conclusie over de verschillen en/of overeenkomsten tussen de beide jaren.

leeftijd	2005 - 2006	2011 - 2012
≤ 13	467	253
14	1095	385
15	1450	503
16	4181	1663
17	10759	3645
18	11465	10883
19	8796	7783
20	6358	5239
21	4632	3533
22	3476	2358
totaal	52679	35245

Tabel 7

Opgave 10

Open het [Excel-bestand Sportprestaties](#). Je vindt er gegevens van 74 leerlingen in de leeftijdscategorie 11 - 12 jaar.

Je wilt de prestaties bij het verspringen van de meisjes en de jongens vergelijken.

- Maak een klassenindeling met een klassenbreedte van 20 cm voor de afstanden van de jongens en de meisjes afzonderlijk. Bereken de relatieve frequenties en maak een staafdiagram.
- Schat vanuit de frequentietabel de modale afstand en het gemiddelde van zowel de jongens als de meisjes.
- Teken boxplots voor de jongens en de meisjes afzonderlijk vanuit de ruwe data. Trek een conclusie.

Toepassen

Je hebt al eerder gewerkt met MS-Excel. Misschien goed om dit weer even te herhalen. Je werkt via het [Practicum](#) met het bestand 'Gegevens154Leerlingen'.

Vervolgens kun je MS-Excel ook goed inzetten bij eigen onderzoekjes. En dat is de beste manier om te leren omgaan met statistiek.

Opgave 11: Cijfer voor wiskunde

Excel kan je helpen bij het berekenen van centrummaten en spreidingsmaten. Bekijk het [Practicum](#).

Denk er wel om dat je op een pc het bijbehorende Excel-bestand eerst moet downloaden.

- Bereken de centrummaten en de spreidingsmaten van de gehele eindcijfers voor wiskunde van deze groep leerlingen in 4vwo.
- Maak van deze eindcijfers een boxplot.
- Om die vraag te kunnen beantwoorden moet je door sorteren eerst de leerlingen die wiskunde A hebben gekozen scheiden van de leerlingen die wiskunde B hebben gekozen. Bereken van beide groepen de centrummaten en de spreidingsmaten van de gehele eindcijfers in 3 vwo. Maak ook twee boxplots.
- Kun je concluderen dat de wiskunde B leerlingen inderdaad betere cijfers voor wiskunde hebben gehaald in 3 vwo?

Opgave 12: Lengtes vergelijken

Werk met hetzelfde Excel-bestand als in de voorgaande opgave. Zet de gegevens over de lengtes van de meisjes en de jongens op een afzonderlijk werkblad.

- Maak een klassenindeling en bereken de bijbehorende centrummaten en spreidingsmaten zowel voor de jongens als de meisjes.
- Probeer een conclusie te trekken uit beide series gegevens.

Opgave 13: Een eigen onderzoek

Bedenk samen met enkele medeleerlingen een onderzoek dat je onder jongeren zou kunnen uitvoeren door middel van een goede representatieve steekproef.

- Omschrijf en leg uit
 - wat je wilt onderzoeken;
 - welke vragen je wilt stellen, of welke gegevens je wilt verzamelen of welke metingen je wilt verrichten;
 - of je denkt dat je steekproef representatief is voor alle jongeren van Nederland;
 - of je denkt dat je steekproef representatief is voor alle 3 vwo leerlingen van Nederland;
 - of je denkt dat je steekproef representatief is voor leerlingen op jouw school.
- Voer je onderzoek uit en verwerk de gegevens met behulp van MS-Excel.
- Probeer conclusies te trekken en die te motiveren met behulp van diagrammen, centrummaten en spreidingsmaten.

Testen

Opgave 14

Deze tabel geeft de eindcijfers voor het vak Duits van twee derde klassen.

- Bereken het gemiddelde eindcijfer van beide klassen.
- Bepaal van beide klassen het modale cijfer.
- Bepaal van beide klassen de mediaan, de kwartielen, de kwartielafstand en de spreidingsbreedte.
- Waarom zijn boxplots nu niet erg nuttig als je beide klassen wilt vergelijken?

eindcijfer Duits	3F	3G
4	2	0
5	3	4
6	6	5
7	5	8
8	5	7
9	2	3
	23	27

Figuur 5

Practicum: Werken met Excel

Excel kan centrummaten en spreidingsmaten voor je berekenen. Je hebt daarvoor geen frequentie-tabellen nodig, die zijn alleen erg handig bij het maken van diagrammen.

In het Excelbestand [Gegevens154Leerlingen.xls](#) vind je enkele gegevens van 154 leerlingen die in 2008 in 4vwo zaten.

Je kunt daarbij van bijvoorbeeld 'cijfwis', het cijfer voor wiskunde van deze leerlingen, het minimum, het eerste kwartiel, de mediaan, het derde kwartiel en het maximum laten berekenen.

Wil je die ook nog in een boxplot weergeven, dan ga je (in Excel2010) zo te werk:

- Maak eerst een kolom met onder elkaar Q1, minimum, mediaan, maximum en Q3 in die volgorde. Geef de kolom ernaast het opschrift 'cijfwis'.
- In de cellen van die kolom maak je de formules =KWARTIEL(H2:H155;1), =MIN(H2:H155), =MEDIAN(H2:H155), =MAX(H2:H155) en =KWARTIEL(H2:H155;3).
- Nu selecteer je beide kolommen en het opschrift en via Invoegen kies je een lijndiagram met markeringen.
- Je krijgt nu een lijndiagram. Klik met de rechtermuisknop op de figuur en kies 'Gegevens selecteren'. Klik daar op 'Schakelen tussen rij en kolom'. Je krijgt dan de vijf waarden boven elkaar.

- Bij 'Hulpmiddelen voor grafieken' kun je bij 'Indeling' kiezen voor 'Omhoog/omlaag balken weer-geven' en 'Hoog/laag lijnen'. Je krijgt dan je boxplot, dat je nog even zo mooi moet maken als je zelf wilt. (Rechtermuisknop op elke markering en opmaak kiezen.)

Misschien wil je de jongens en de meisjes vergelijken? Dan laat je Excel eerst sorteren op de kolom 'geslacht'. Dat doe je door die kolom te selecteren en 'Sorteren en filteren' te kiezen. Sorteert van laag naar hoog en klik op OK als Excel je vraagt of de selectie moet worden uitgebreid (want de andere kolommen moeten ook worden meegenomen bij het sorteren). Nu heb je de jongens en de meisjes als groepen bij elkaar staan en kun je hun centrummaten en spreidingsmaten vergelijken en twee boxplots in één figuur maken.



© 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.
