



Het gemiddelde

Prof. dr. Herman Callaert

Inhoudstafel

1	Het intuïtieve begrip “gemiddelde”	1
1.1	Er samen voor opdraaien	1
1.2	Eerlijk verdelen	1
2	Spelen met de bouwstenen van het gemiddelde	2
2.1	Drie fundamenten	2
2.2	Ken je er twee dan weet je er drie!	3
2.3	Elk afzonderlijk of toch weer niet?	4
2.4	Groter en kleiner, maar samen nul	5
2.5	Neem me maar niet te letterlijk!	6
3	Cijfermateriaal samenvatten en het gemiddelde	7
3.1	Voorbeeld 1: schoolboeken	7
3.2	Voorbeeld 2: schoolboeken (vervolg)	9
3.3	Voorbeeld 3: diameters	10
4	Bewerkingen op getallen en de invloed daarvan op het gemiddelde	12
4.1	Voorbeeld 4: schoolboeken en klasagenda’s	12
4.2	Voorbeeld 5: systematische meetfout	14
4.3	Voorbeeld 6: lengte van kinderen	14

1 Het intuïtieve begrip “gemiddelde”

1.1 Er samen voor opdraaien

In de geschiedenis kan je heel wat voorbeelden vinden waarbij een of ander losweg gedefinieerd begrip van gemiddelde te bespeuren valt. Oorspronkelijk ging dat nog niet over de exacte formule van het rekenkundig gemiddelde, maar over “representatief zijn” of over “kosten delen”. Zo was er reeds eeuwen vóór Christus een bloeiende handel tussen steden aan de Middellandse zee, en veel koopwaar werd per boot over zee vervoerd. Dergelijke transporten waren niet zonder risico, en bij storm was er soms veel schade aan de schepen. Ook gebeurde het dat de kapitein verplicht was om koopwaar over boord te gooien opdat het zwaarbeladen schip niet helemaal in de storm zou vergaan.

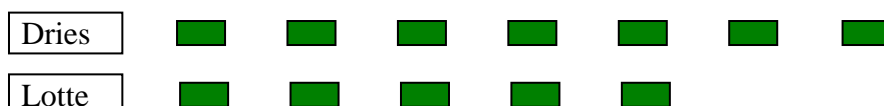
Na zo’n onfortuinlijke tocht was er heel wat averij. Meestal vervoerde het schip koopwaar van meerdere handelaars, en het is nogal duidelijk dat de kapitein bij een storm geen tijd had om ervoor te zorgen dat hij van elke handelaar evenveel overboord gooide. Daarom waren er afspraken tussen de handelaars. Bij tegenslag (kapers, storm, ...) zou iedereen een “eerlijke” bijdrage leveren. Het bedrag van die bijdrage werd bepaald door ingewikkelde regels waarbij de waarde van de beschadigde lading moest worden geschat, de averij aan het schip, enz..

In het Engels zijn er twee woorden voor gemiddelde: “mean”, waarbij het rekenkundig gemiddelde wordt bedoeld, en “average”, dat zowel op het rekenkundig gemiddelde wijst als op de meer intuïtieve benadering zoals boven beschreven. Sommigen denken zelfs dat “average” afkomstig is van het Arabische “awariyah” (beschadigde goederen). Zou ons woord “averij” daar ook iets mee te maken hebben?

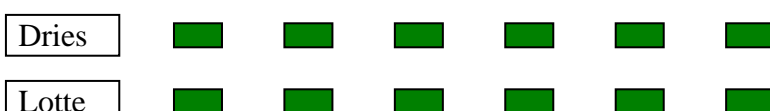
1.2 Eerlijk verdelen

Het zit de mens blijkbaar ingebakken om zichzelf te vergelijken met anderen. Bij kinderen speelt dikwijls ook een spontaan “rechtvaardigheidsgevoel” mee, zeker als Lotte minder snoepjes gekregen heeft dan Dries. Dat is niet eerlijk, vindt Lotte, en dus moet er “herverdeeld” worden. Dit kan op allerlei manieren gebeuren, zelfs zonder tellen. Zij kunnen bijvoorbeeld twee rijen snoepjes maken door ze naast elkaar te leggen, en dan snoepjes van de langste rij naar de kortste verplaatsen tot beide rijen even lang zijn. (Bij een oneven aantal snoepjes zal papa of mama moeten ter hulp komen!)

De oorspronkelijke situatie



wordt “eerlijkheidshalve” herstampeld tot



Iets moeilijker wordt het wanneer je repen chocolade hebt die niet allemaal even lang zijn. Dan heb je een mes nodig, en misschien kunnen kinderen dat dan toch maar beter aan mama vragen.



Mama ziet dat Lies weer het grootste stuk genomen heeft, en dus zal zij daar wat van afsnijden om aan de anderen te geven. Maar zo simpel is het niet, want als er bij Lies te veel chocolade wordt afgesneden, dan heeft Marc meer dan de anderen. Om daar een beter zicht op te krijgen legt mama de chocoladerepen in een andere volgorde. Haar geoefende oog heeft snel gezien dat zij de chocolade die Cecile en Hans te weinig heeft moet compenseren met wat Lies en Marc te veel heeft.



Zowel bij de snoepjes als bij de chocolade wordt er gezocht naar een nieuwe plaats (een nieuw aantal snoepjes of een nieuwe lengte van de chocoladerepen). Die nieuwe plaats wordt zo bepaald dat wat de enen te veel hebben exact overeenkomt met wat de anderen te weinig hebben. Als je dan herverdeelt, dan heeft iedereen juist evenveel, namelijk zoveel als wordt aangegeven door die nieuwe plaats. En bovendien zijn er geen snoepjes of stukjes chocolade verdwenen. De kinderen, die vinden dat alles “eerlijk” moet verdeeld worden, zijn tevreden.

2 Spelen met de bouwstenen van het gemiddelde

2.1 Drie fundamenteën

De chocoladereep van Lies was 9 cm lang. Dit is de eerste chocoladereep die je tegenkomt, en haar lengte (in cm) noteer je door de letter x met index één, dus x_1 . Hans had maar een heel klein stuk chocolade, dat was maar 3 cm lang. Het getal 3 noteer je als x_2 . Bij Marc en Cecile was het 7 cm en 5 cm. Alles samen heb je dus dat $x_1 = 9$, $x_2 = 3$, $x_3 = 7$, en $x_4 = 5$.

De totale lengte van al die chocoladerepen is $x_1 + x_2 + x_3 + x_4 = 9 + 3 + 7 + 5 = 24$ cm,

wat je ook kan schrijven als $\sum_{i=1}^4 x_i = x_1 + x_2 + x_3 + x_4 = 24$ cm.

Het was de taak van mama om elk kind evenveel chocolade te geven. Eigenlijk was dat een dubbele opdracht. Er mocht geen enkel kind vergeten worden, en elk kind moest evenveel krijgen. Het aantal kinderen noteer je door n en hoeveel chocolade ze krijgen (in cm) noteer je door \bar{x} .

In dit voorbeeld is $n = 4$, maar wat is \bar{x} ? Als elk van de 4 kinderen \bar{x} cm chocolade krijgt, dan hebben zij samen $4 \times \bar{x}$ cm chocolade. Er is geen chocolade bijgekomen en er is er ook geen verdwenen, zij is gewoon “herverdeeld”. En alle chocolade samen, dat was $\sum_{i=1}^4 x_i$ cm.

Dit betekent dat $4 \times \bar{x}$ cm = $\sum_{i=1}^4 x_i$ cm of dat $4 \times \bar{x} = 24$. Dus is $\bar{x} = 6$. Elk kind heeft nu evenveel chocolade, namelijk 6 cm. Het getal dat ervoor zorgt dat je “aan iedereen evenveel” kan geven noteer je door \bar{x} en je noemt dat “het gemiddelde”.

De drie grootheden die hier met elkaar verbonden zijn zien er als volgt uit:

- het **aantal** getallen x_i . Hier zijn er dat 4, maar algemeen noteer je het aantal door n
- de **som** van die n getallen, namelijk $x_1 + x_2 + \dots + x_n$, wat je op een korte manier opschrijft als $\sum_{i=1}^n x_i$
- het **gemiddelde** van die n getallen, genoteerd door \bar{x} .

2.2 Ken je er twee dan weet je er drie!

Het is voldoende om twee van de drie grootheden te kennen om ook de derde te weten te komen. Probeer dat maar even uit op het chocoladevoorbeeld (waarbij overal de lengte in cm is opgeschreven).

Als je weet dat de som 24 is ($\sum_{i=1}^4 x_i = 24$) en dat het aantal kinderen 4 is ($n = 4$), dan weet je

ook wat het gemiddelde \bar{x} is, want uit $4 \times \bar{x} = \sum_{i=1}^4 x_i$ volgt dat $\bar{x} = \frac{1}{4} \sum_{i=1}^4 x_i$ zodat

$$\bar{x} = \frac{24}{4} = 6.$$

Wanneer je voor het “aantal” de algemene notatie n gebruikt, dan heb je dat $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Dit kan je in woorden lezen als: “het gemiddelde van getallen is gelijk aan de som van die getallen gedeeld door het aantal getallen”.

Onderstel eens dat je weet dat het gemiddelde \bar{x} gelijk is aan 6 en dat er 4 kinderen zijn ($n = 4$), wat is dan de som $\sum_{i=1}^4 x_i$ van alle getallen? Dat heb je zopas berekend, en uit

$4 \times \bar{x} = \sum_{i=1}^4 x_i$ volgt dat $\sum_{i=1}^4 x_i = 4 \times 6 = 24$. In algemene notatie schrijf je dit als

$\sum_{i=1}^n x_i = n \bar{x}$. In woorden zeg je: “de som van alle getallen is gelijk aan n keer het gemiddelde, waarbij n het aantal getallen is”.

Onderstel tenslotte eens dat je weet dat de som van de lengte (in cm) van al die chocoladerepen gelijk is aan 24. Na “eerlijk herverdelen” heeft elk kind een stuk gekregen met lengte (in cm) van $\bar{x} = 6$. Hoeveel kinderen zijn er dan wel (wat is n)?

Ook hier kan je op antwoorden. Elk van de n kinderen heeft een stuk gekregen van lengte \bar{x} , wat een totaal oplevert van $n \times \bar{x} = n \times 6$. Maar anderzijds was de som $\sum_{i=1}^n x_i$ gelijk aan

24. Dus moet $n \times \bar{x} = \sum_{i=1}^n x_i$, zodat, na beide leden te delen door \bar{x} ,

$$n = \frac{\sum_{i=1}^n x_i}{\bar{x}}$$

Je kan dit als volgt zeggen: “het aantal getallen is gelijk aan hun som gedeeld door hun gemiddelde”.

In ons voorbeeld zijn er dus $n = \frac{\sum_{i=1}^n x_i}{\bar{x}} = \frac{24}{6} = 4$ kinderen.

2.3 Elk afzonderlijk of toch weer niet?

Als je alle afzonderlijke getallen kent, dan weet je ook hoeveel er zijn, en bovendien kan je dan hun som berekenen. Dat is voldoende om te weten wat het gemiddelde van die getallen is,

want $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Maar heb je alle getallen echt nodig om het gemiddelde te berekenen?

Herinner je dat: “het gemiddelde van getallen is gelijk aan de som van die getallen gedeeld door het aantal getallen”. Als je dus enkel weet dat alle chocoladerepen samen een lengte opleveren van 24 cm en dat er 4 kinderen zijn dan weet je dat elk kind 6 cm zal krijgen.

Je komt hier tot de ontdekking dat, om het gemiddelde \bar{x} van n getallen $\{x_1, x_2, \dots, x_n\}$ te berekenen, je niet noodzakelijk die getallen moet kennen. Als iemand je zegt dat het over n getallen gaat en je daarbij ook vertelt hoeveel hun som $\sum_{i=1}^n x_i$ is, dan weet je genoeg.

Omgekeerd, als iemand je zegt over hoeveel getallen het gaat (wat n is), en wat hun gemiddelde \bar{x} is, kan je dan die getallen $\{x_1, x_2, \dots, x_n\}$ zelf ook te weten komen? Neen!

Als je weet dat 4 kinderen na herverdelen elk een chocoladereep van 6 cm hebben, dan ken je n en \bar{x} , maar je weet helemaal niet hoeveel Lies, Hans, Marc en Cecile oorspronkelijk hadden. In ons voorbeeld was dat (in cm): 9, 3, 7, en 5. Maar je kan eenvoudig natellen dat het ook helemaal anders had kunnen zijn, zoals bijvoorbeeld 12 cm voor Lies, en slechts 4 cm voor ieder van de drie anderen. De informatie die in het gemiddelde \bar{x} van n getallen zit laat je niet toe om die getallen $\{x_1, x_2, \dots, x_n\}$ terug te vinden.

2.4 Groter en kleiner, maar samen nul

Bij het herverdelen van de chocoladerepen is er gezocht naar een lengte \bar{x} (in dit voorbeeld was dat 6 cm) waarvoor het volgende waar is. Lies heeft $x_1 = 9$ cm, en dat is meer dan $\bar{x} = 6$ cm. Wat zij meer heeft is $x_1 - \bar{x} = 3$ cm. Marc heeft ook meer en voor hem is dat $x_3 - \bar{x} = 1$ cm. Maar Hans en Cecile hebben minder dan het gemiddelde, en als je voor hen dezelfde berekening maakt dan vind je (in cm) $x_2 - \bar{x} = 3 - 6 = -3$ en $x_4 - \bar{x} = 5 - 6 = -1$. Je ziet hier getallen x_i waarvoor het verschil $x_i - \bar{x}$ positief is, maar je hebt ook getallen x_i waarvoor het verschil $x_i - \bar{x}$ negatief is. Als je alle positieve verschillen samentelt dan is dat (op het teken na) exact gelijk aan de som van al de negatieve verschillen. Dit betekent dat de totale som van al die verschillen gelijk is aan nul.

Voor elke verzameling getallen $\{x_1, x_2, \dots, x_n\}$
 is het altijd waar dat $\sum_{i=1}^n (x_i - \bar{x}) = 0$

Deze eigenschap kan je eenvoudig aantonen als je je herinnert dat $\sum_{i=1}^n x_i = n \bar{x}$ en dat

$\sum_{i=1}^n \bar{x} = \bar{x} + \bar{x} + \dots + \bar{x} = n \bar{x}$. Je hebt dan immers

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = n \bar{x} - n \bar{x} = 0$$

2.5 *Neem me maar niet te letterlijk!*

Als je n getallen $\{x_1, x_2, \dots, x_n\}$ hebt, dan kan je daarvan het gemiddelde \bar{x} berekenen. Dat gemiddelde is terug een getal. Is dat gelijk aan één van de oorspronkelijke getallen? Niet noodzakelijk. In het voorbeeld van de chocolade was het gemiddelde 6 cm, en er was geen enkel kind dat in het begin een reep van 6 cm genomen had. Maar je had hier wel geluk dat de totale lengte zo mooi te verdelen was onder die 4 kinderen.

Onderstel eens dat er slechts 3 kinderen zijn, met chocoladerepen waarvan de lengte gelijk is aan 8 cm, 4 cm, en 10 cm. Het gemiddelde hiervan is $\bar{x} = \frac{1}{3} \sum_{i=1}^3 x_i = \frac{1}{3} (8 + 4 + 10) = \frac{22}{3}$ cm. Hier heb je een probleem, want 22 gedeeld door 3 geeft iets van de vorm 7.3333333... .

Een klassieke afspraak is dat je, waar nodig, het gemiddelde opschrijft met één decimale plaats meer dan de oorspronkelijke getallen. In dit voorbeeld waren de oorspronkelijke getallen (in cm) gelijk aan 8, 4, en 10. Voor het gemiddelde schrijf je dan dat \bar{x} (in cm) gelijk is aan 7.3 .

De moeilijkheid die je hier tegenkomt is niet alleen een wiskundig probleem. Zij zorgt ook voor een interpretatieprobleem. Je moet hierbij goed in het oog houden dat het gemiddelde, als nieuw getal, ontstaan is door een wiskundige bewerking uit te voeren (de som delen door het aantal). En wiskundige bewerkingen storen zich niet aan de context. Als je dat weet, dan weet je ook hoe je een gemiddelde moet interpreteren.

Wat dacht je anders van de uitspraak: “in ons land heeft een gezin gemiddeld 2.1 kinderen”? Waarschijnlijk denk jij daar anders over dan een kind uit de lagere school in Australië dat zei: “dat betekent dat elk gezin twee kinderen heeft, en dat de mama terug in verwachting is”!

3 Cijfermateriaal samenvatten en het gemiddelde

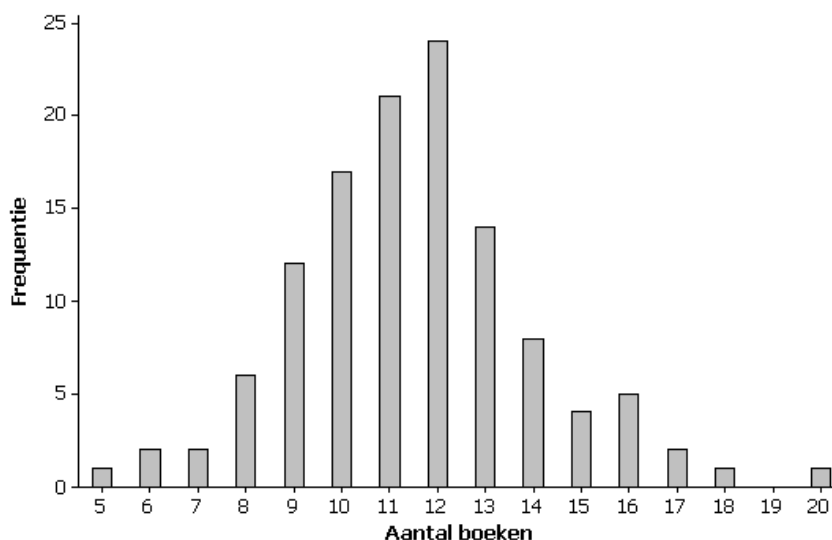
Bij numerieke opmetingen bepaal je eerst of je methoden voor categorische data zal gebruiken of methoden voor continue data. Zodra je dat weet kan je een aangepaste figuur maken, en dat is het eerste wat je moet doen bij je exploratie van een dataset. Een figuur kan je immers helpen ontdekken of je het cijfermateriaal zinvol kan samenvatten in enkele kengetallen, en of hierbij het gebruik van het gemiddelde een goede keuze is.

3.1 Voorbeeld 1: schoolboeken

Op een donderdag werd in een school aan 120 leerlingen gevraagd hoeveel boeken (handboeken, schriften en ringmappen) zij die dag hadden meegebracht. Het antwoord was als volgt.

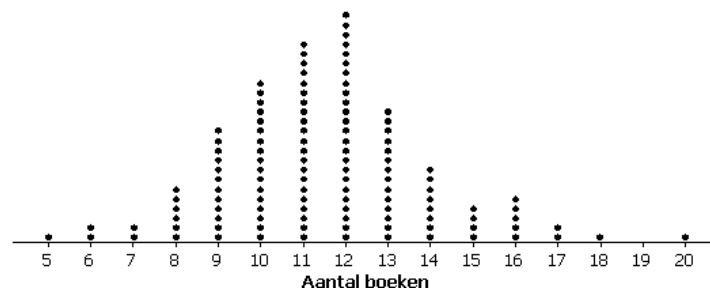
12	13	11	13	13	12	17	20	13	12	12	12
11	14	10	11	9	10	9	13	7	9	11	12
12	12	12	11	10	9	7	9	11	11	13	11
12	9	10	16	16	14	8	11	10	11	13	8
13	12	15	12	13	10	17	16	12	12	16	14
11	15	9	11	12	12	6	8	12	12	11	10
16	5	9	8	10	15	10	8	9	12	13	12
11	11	11	18	11	11	12	12	12	13	9	8
10	13	12	14	14	11	10	11	10	9	10	15
9	10	6	11	10	13	14	14	14	10	10	13

Het gaat hier over “het aantal boeken”, wat een discrete veranderlijke is met een beperkt aantal verschillende uitkomsten (het minimum is 5 en het maximum is 20 in dit voorbeeld). Een aangepaste grafiek hiervoor is het staafdiagram.

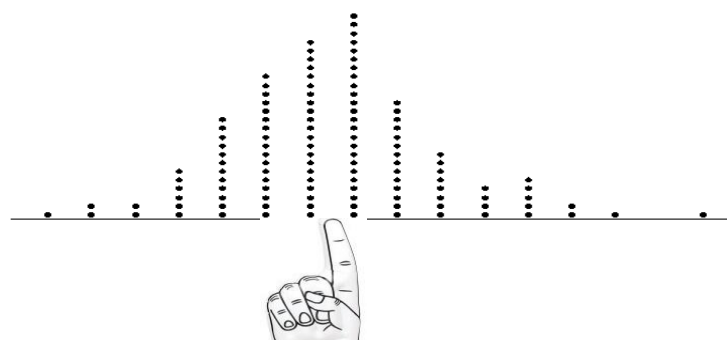


Als je naar de “globale” vorm van de figuur kijkt, dan bemerk je dat er geen opvallende “pieken of gaten” in zitten. Er is natuurlijk wat variabiliteit, en helemaal symmetrisch is hij ook niet. Maar dat is niet erg. Het “globale” patroon zegt dat er heel wat waarnemingsgetallen in het centrum liggen (ergens tussen 10 en 13) en dat er minder en minder waarnemingsgetallen voorkomen naarmate je je verder van dit centrum verwijderd (zowel naar links als naar rechts). In een dergelijke situatie kan je zeggen dat het gemiddelde “typisch” is voor die opmetingen. In dit voorbeeld is het gemiddelde dus “typisch” voor het aantal boeken dat die 120 leerlingen in die school op die dag bij zich hadden. Het gemiddelde van die 120 waarnemingsgetallen is $\bar{x} = 11.5$ en je kan zeggen dat een “typische” leerling zo’n 11 à 12 boeken bij heeft. Daarbij verwacht je ook heel wat leerlingen te vinden met 10 of 13 boeken, maar veel minder leerlingen met een aantal boeken dat ver van het gemiddelde ligt, zoals 6 of 18.

Het gemiddelde heeft ook een mooie “fysische” interpretatie. Daarbij doe je het volgende (denkbeeldige) experiment. Neem de getallenas, en denk dat die eruit ziet als een lange gewichtsloze staaf. Plaats op die getallenas een gewichtje (van bijvoorbeeld 1 gram) op elke plaats waar je waarnemingsgetallen hebt. Je plaatst dus één gewichtje op 5, want er was één leerling met 5 boeken. Er waren 2 leerlingen met 6 boeken, en dus plaatst je twee gewichtjes op het getal 6, enz.. Een grafiek die dit mooi illustreert is het puntendiagram.



Wel, nu kan je “voelen” waar het gemiddelde ligt. Zet de figuur op je vinger op die plaats waar zij in evenwicht blijft staan. Daar ligt het gemiddelde. In de fysica zou je dit het “evenwichtspunt” noemen.

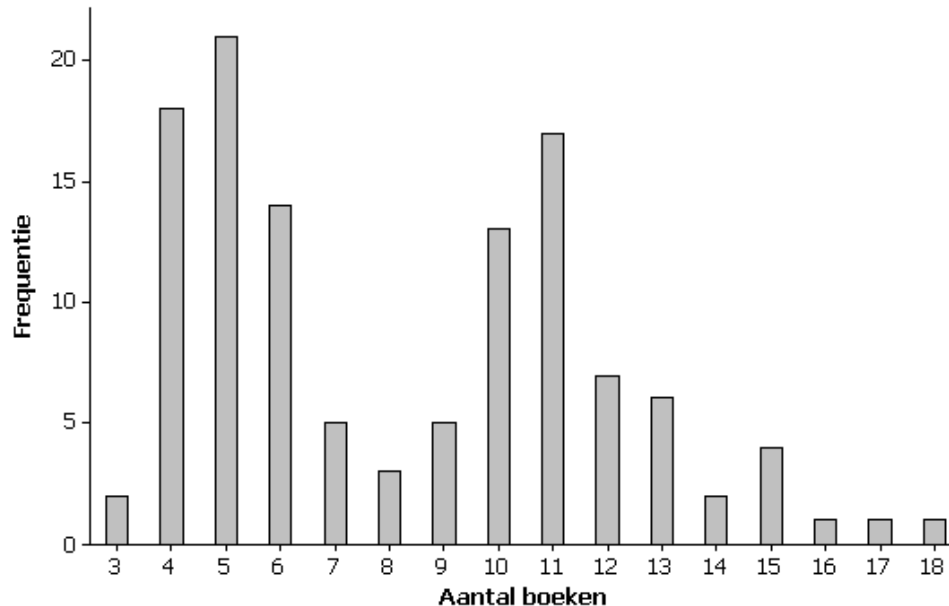


3.2 Voorbeeld 2: schoolboeken (vervolg)

In een andere school wil men ook eens weten hoeveel boeken hun leerlingen dagelijks meesleuren. Twee leerkrachten (die niet in dezelfde klassen komen) spreken af dat zij dat elk aan 60 leerlingen zullen vragen. Zij zullen daarna hun getallen samen leggen. Zo hebben zij ook 120 opmetingen, en die staan hieronder.

8	4	5	10	11	12	16	5	5	11	6	14
4	5	11	17	6	5	10	14	6	4	13	12
6	4	7	10	11	6	11	12	5	3	10	9
4	11	6	11	8	13	11	8	5	10	10	11
4	18	7	10	10	5	10	10	6	11	9	11
10	7	11	5	5	5	5	15	4	6	7	4
5	9	13	9	5	5	6	7	4	5	15	5
15	4	13	4	4	6	3	4	4	11	9	12
13	11	4	12	10	10	11	5	5	5	6	4
12	11	13	15	6	6	4	12	6	5	11	4

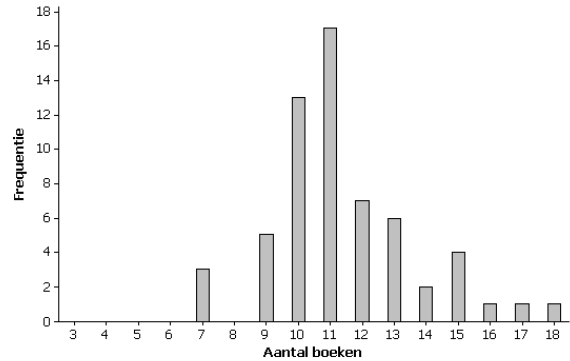
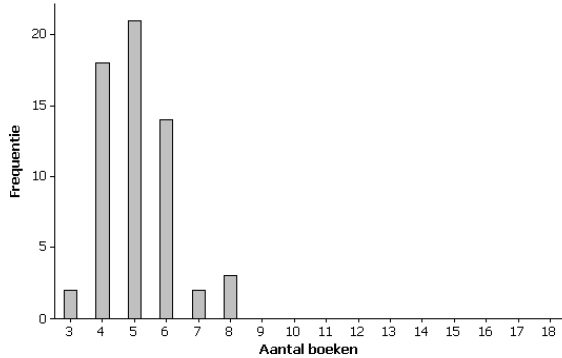
Het gemiddelde van deze 120 getallen is $\bar{x} = 8.2$. Is dit gemiddelde “typisch” voor deze 120 leerlingen? Is het waar dat de meerderheid van die leerlingen rond de acht boeken per dag meebrengt? Om dit te weten te komen heb je niet genoeg aan één enkel getal, zoals het gemiddelde. Een figuur brengt meestal heel wat extra inzicht. Bekijk maar eens dat staafdiagram.



Het blijkt helemaal niet waar te zijn dat een groot aantal leerlingen “ongeveer” 8 boeken bij heeft. Het staafdiagram vertelt een heel ander verhaal. Er blijken 2 clusters te zijn, één rond 5 en één rond 11. Hoe kan dat verklaard worden?

Misschien kom je meer te weten als je eens vraagt hoe die getallen tot stand zijn gekomen. En inderdaad, in dit geval was de oorzaak niet moeilijk te achterhalen. Eén van de leerkrachten had die vraag gesteld op een donderdag maar de andere leerkracht had die vraag gesteld op

een woensdag, en dan is er maar een halve dag les. De staafdiagrammen voor die twee groepen (elk gebaseerd op 60 opmetingen) zien er als volgt uit.



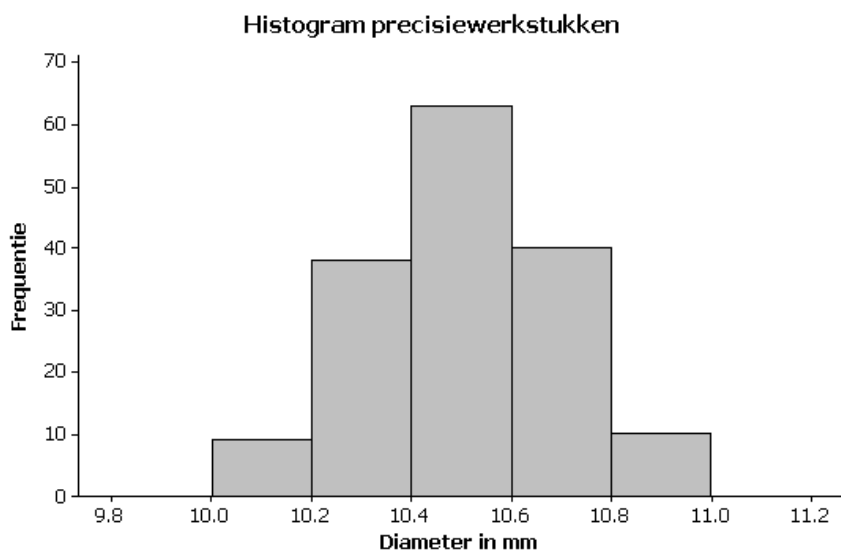
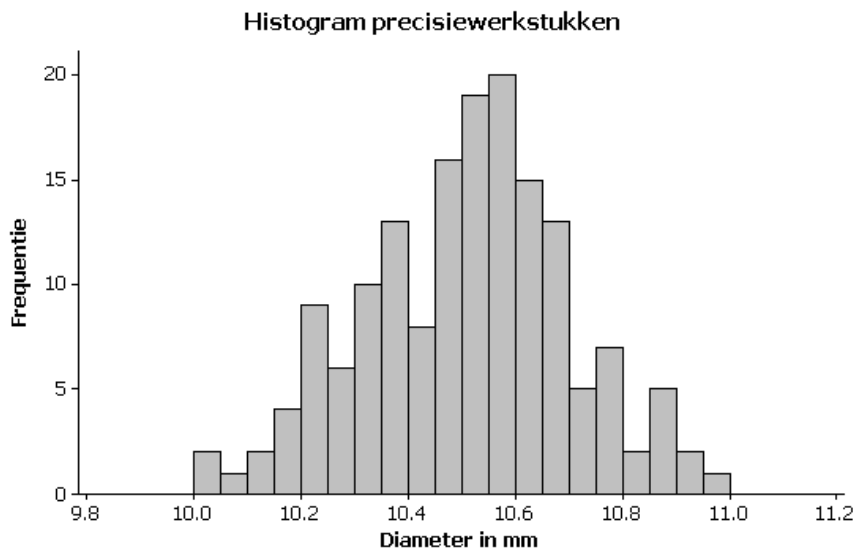
Voor de eerste groep is het gemiddelde 5.1, en voor de tweede groep is dat 11.4. Beide staafdiagrammen vertonen nu geen eigenaardige patronen meer, en je kan zeggen dat het gemiddelde $\bar{x} = 5.1$ “typisch” is voor die eerste groep. Die 60 leerlingen die over de middag thuis gaan eten hebben per halve dag ongeveer 5 boeken bij. Die andere 60 leerlingen brengen ongeveer 11 boeken mee voor een volledige schooldag. Maar het gemiddelde van alle getallen samen, namelijk $\bar{x} = 8.2$, is in dit voorbeeld geen goede maat voor het “typisch” aantal boeken van die 120 leerlingen.

3.3 Voorbeeld 3: diameters

Hieronder zie je de diameter (opgemeten in mm) van precisiewerkstukken. Het is de bedoeling werkstukken te produceren die een diameter van 10.50 mm hebben, maar een klein verschil is aanvaardbaar.

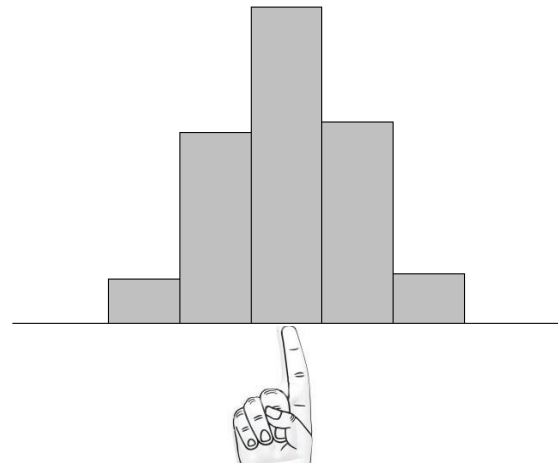
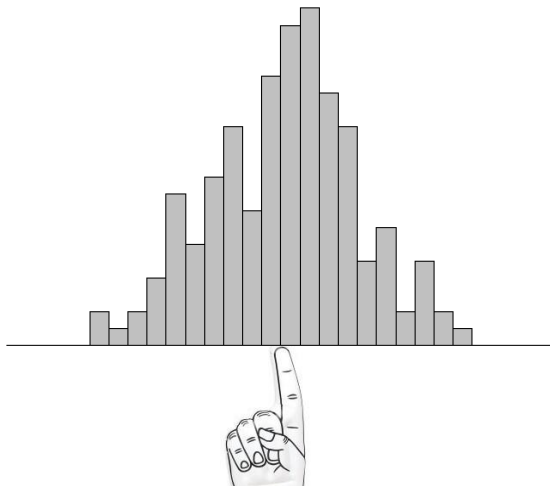
10.59	10.48	10.48	10.97	10.84	10.54	10.39	10.33	10.55	10.47	10.15	10.30
10.78	10.63	10.58	10.25	10.86	10.45	10.27	10.68	10.42	10.69	10.36	10.66
10.49	10.28	10.34	10.66	10.61	10.56	10.85	10.62	10.43	10.53	10.43	10.44
10.60	10.38	10.45	10.31	10.72	10.40	10.37	10.04	10.22	10.70	10.58	10.79
10.66	10.22	10.57	10.52	10.55	10.23	10.47	10.53	10.53	10.53	10.53	10.58
10.69	10.30	10.35	10.23	10.47	10.88	10.52	10.20	10.56	10.52	10.72	10.62
10.65	10.46	10.52	10.25	10.45	10.30	10.39	10.58	10.18	10.21	10.66	10.39
10.82	10.85	10.22	10.50	10.33	10.58	10.62	10.59	10.63	10.41	10.36	10.77
10.64	10.48	10.72	10.21	10.52	10.78	10.54	10.05	10.76	10.55	10.58	10.51
10.39	10.47	10.90	10.13	10.61	10.71	10.32	10.27	10.60	10.68	10.52	10.75
10.15	10.33	10.75	10.42	10.16	10.57	10.51	10.37	10.64	10.65	10.45	10.36
10.51	10.47	10.24	10.64	10.56	10.45	10.58	10.55	10.54	10.49	10.54	10.64
10.65	10.38	10.12	10.67	10.58	10.30	10.60	10.37	10.42	10.60	10.87	10.93
10.55	10.01	10.27	10.68								

Je hebt hier te maken met opgemeten lengten, en dat is een grootte die je als continu behandelt. Een aangepaste figuur hiervoor is het histogram. Afhankelijk van de keuze van de klassenbreedte krijg je verschillende voorstellingen van eenzelfde dataset. Hieronder zie je twee histogrammen voor die diameters. Het histogram met de kleinere klassenbreedte laat wat meer variabiliteit zien, maar in beide gevallen zie je in de “globale” figuur geen onverwachte patronen. Dit is dus terug een situatie waarbij het gemiddelde je een goed idee geeft over de waarde van een “typische” diameter van deze 160 werkstukken. Het gemiddelde is hier gelijk aan $\bar{x} = 10.503$ mm. Dit weet je omdat je over de oorspronkelijke gegevens beschikt. Als je alleen een histogram (of frequentietabel) hebt, dan vind je een benaderende waarde voor \bar{x} .



Ook in het continue geval kan je de “fysische” interpretatie van het gemiddelde illustreren. Maar nu zal je geen gewichtjes meer plaatsen op bepaalde punten. Bij een continuüm wordt het aantal waarnemingsgetallen dat in een interval terecht komt voorgesteld (op een evenredigheidsfactor na) door de oppervlakte boven dat interval. Dit is tenandere de basisregel voor het tekenen van histogrammen.

Je kan je nu voorstellen dat die histogram bestaat uit één of andere uniforme vaste stof. Denk gewoon aan een stuk karton dat je kan uitknippen. En dan kan je weer het gemiddelde “voelen” door naar het evenwichtspunt te zoeken.

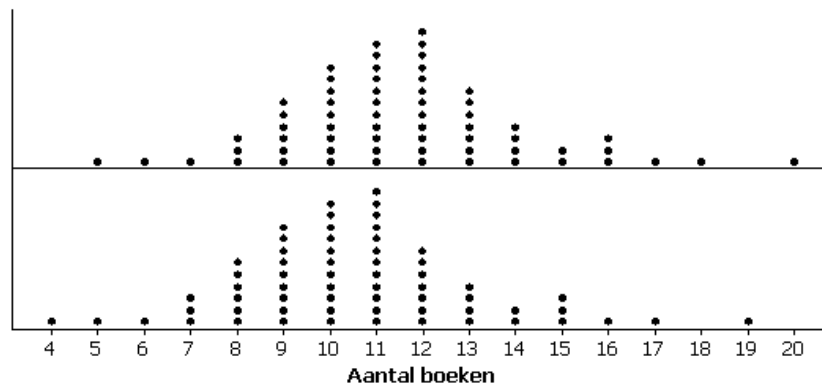


4 Bewerkingen op getallen en de invloed daarvan op het gemiddelde

4.1 Voorbeeld 4: schoolboeken en klasagenda's

In voorbeeld 1. hebben de leerlingen alle boeken geteld die zij in hun boekentas hadden. Maar eigenlijk wou de directie te weten komen hoeveel studieboeken zij naast hun klasagenda bij hadden. Het getal dat elke leerlingen heeft opgegeven is dus één eenheid te groot.

Je hebt hier eigenlijk te maken met de volgende situatie. Je vertrekt met $n = 120$ oorspronkelijke getallen x_i waarvan je het gemiddelde $\bar{x} = 11.5$ hebt berekend. Maar nu moet je overstappen op nieuwe getallen y_i waarbij $y_i = x_i - 1$ want van elke x_i moet een vast getal 1 worden afgetrokken. Je zou nu expliciet die 120 nieuwe getallen y_i kunnen uitrekenen, en dan zou je vinden: $y_1 = x_1 - 1 = 12 - 1 = 11$, $y_2 = x_2 - 1 = 13 - 1 = 12$, enz.. Maar moet je al die nieuwe getallen y_i hebben vooraleer je hun gemiddelde \bar{y} kan bepalen? Kijk eens grafisch wat er gebeurt. Teken een staafdiagram (of een puntendiagram). Wat zie je? En hoe kan je nu het nieuwe gemiddelde \bar{y} “voelen”? In welke richting moet je je vinger verschuiven, en over welke afstand?



Je bemerkt dat het bovenste puntendiagram als een “bevroren geheel” naar links verschoven is over een afstand van één eenheid. Het “evenwichtspunt” van het nieuwe onderste puntendiagram vind je dus ook één eenheid lager, namelijk in het punt 10.5.

De eigenschap die je hier grafisch hebt ontdekt is algemeen waar, en je kan die als volgt formuleren. Start met n getallen x_i waarvan het gemiddelde gelijk is aan \bar{x} . Als je bij elk getal x_i een vast getal a optelt (of ervan aftrekt) dan krijg je nieuwe getallen y_i . Wanneer je nu alleen maar het gemiddelde \bar{y} van die nieuwe getallen moet weten, dan hoef je niet eerst al die y_i uit te rekenen. Het nieuwe gemiddelde \bar{y} vindt je immers door bij het oude gemiddelde \bar{x} het vaste getal a op te tellen (of ervan af te trekken).

Voor elk vast getal a (zowel positief als negatief) geldt:

$$y_i = x_i + a \quad (\text{voor } i = 1, 2, \dots, n) \Rightarrow \bar{y} = \bar{x} + a$$

Als je deze eigenschap echt wilt bewijzen, dan kan je handig gebruik maken van de bouwstenen van het gemiddelde. Om van n getallen y_i het gemiddelde \bar{y} te zoeken, heb je hun som nodig. Maar je weet dat $y_i = x_i + a$ en dus is

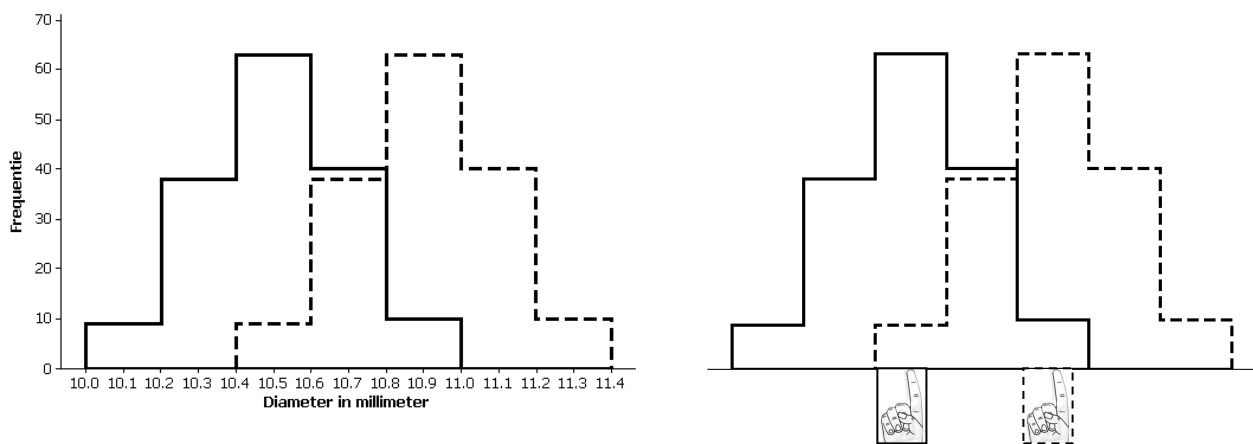
$$\begin{aligned} \sum_{i=1}^n y_i &= \sum_{i=1}^n (x_i + a) = (x_1 + a) + (x_2 + a) + \dots + (x_n + a) \\ &= (x_1 + x_2 + \dots + x_n) + (a + a + \dots + a) = \left(\sum_{i=1}^n x_i \right) + (n \times a) \end{aligned}$$

Nu kan je de som delen door het aantal, en dan heb je het gemiddelde. Dat wordt:

$$\bar{y} = \frac{1}{n} \left[\left(\sum_{i=1}^n x_i \right) + (n \times a) \right] = \frac{1}{n} \left(\sum_{i=1}^n x_i \right) + \frac{1}{n} (n \times a) = \bar{x} + a, \text{ en dus is } \bar{y} = \bar{x} + a.$$

4.2 Voorbeeld 5: systematische meetfout

Indien voor het meten van die 160 diameters het elektronisch meettoestel systematisch vier tienden van een millimeter te weinig zou geregistreerd hebben, dan zou je moeten overstappen van de oude x_i getallen naar nieuwe y_i getallen, met $y_i = x_i + 0.4$. Je kan hier terug de aangepaste grafieken tekenen, en je bemerkt dat het oorspronkelijke histogram (volle lijn) over een afstand van 0.4 eenheden naar rechts verschoven is (stippelijjn). Het “evenwichtspunt van het nieuwe histogram, en dus ook het gemiddelde \bar{y} van de nieuwe y_i getallen, is met 0.4 mm vergroot. Je weet dat $\bar{x} = 10.503$ mm, en dus vind je eenvoudig dat $\bar{y} = \bar{x} + 0.4 = 10.503 + 0.4 = 10.903$ mm. Als het waar is dat die 160 precisiewerkstukken een diameter van 10.50 mm moeten hebben, en dat er slechts een lichte afwijking mag zijn, dan zou je nu wel eens een probleem kunnen hebben.

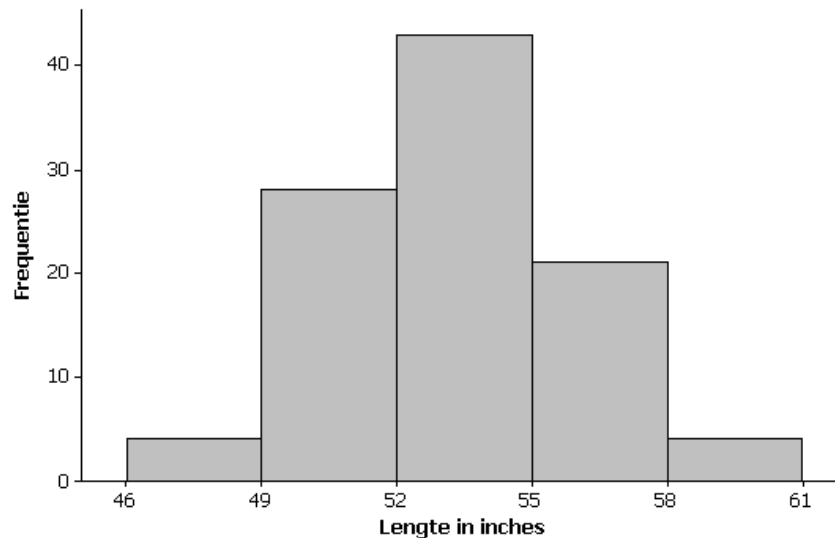


4.3 Voorbeeld 6: lengte van kinderen

In Californië werd in de jaren zeventig een onderzoek uitgevoerd in het kader van de volksgezondheid. Daarbij werden heel wat gegevens genoteerd bij de geboorte van een kind, en ook bij de test die 10 jaar later plaatsvond. Eén van de opgemeten veranderlijken was de lengte van het kind van 10 jaar. Hieronder staan zo'n 100 opmetingen.

Als eerste getal zie je daar 50.5 staan, en ook alle andere getallen zijn van dezelfde grootteorde. Dit doet je de vraag stellen in welke eenheid die lengten wel zijn opgemeten. En inderdaad, als je in Californië aan een kind van 10 jaar vraagt: “hoe groot ben je?” dan krijg je kort en bondig “I’m four two”, wat voluit wil zeggen “I’m four feet and two inches”. Als je alles omrekent naar inches (1 foot = 12 inches) krijg je dat dit kind 50 inches groot is. En dit is inderdaad de eenheid waarin al die lengten zijn genoteerd.

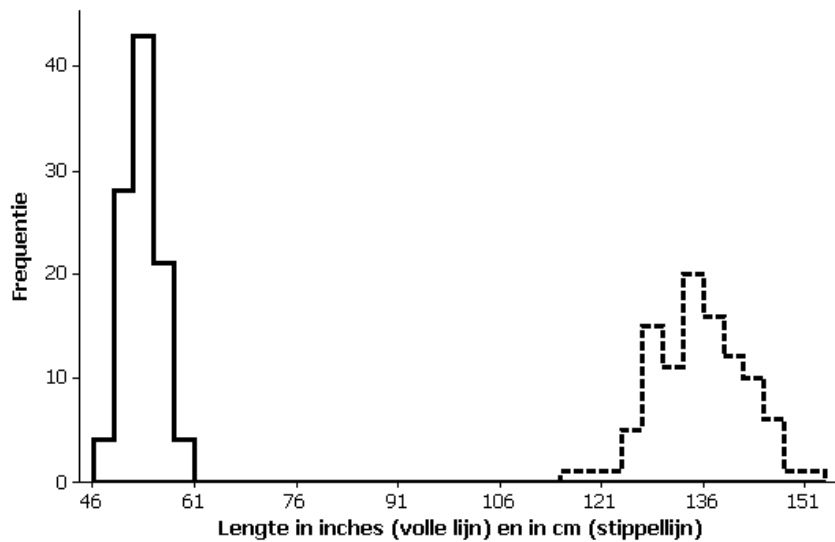
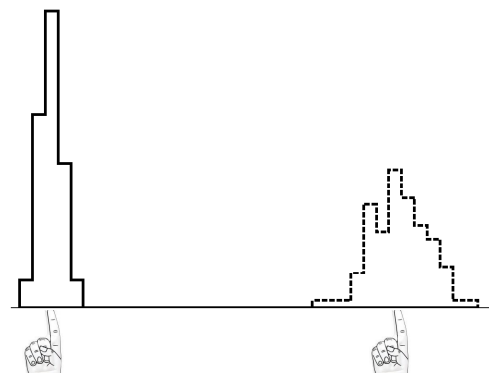
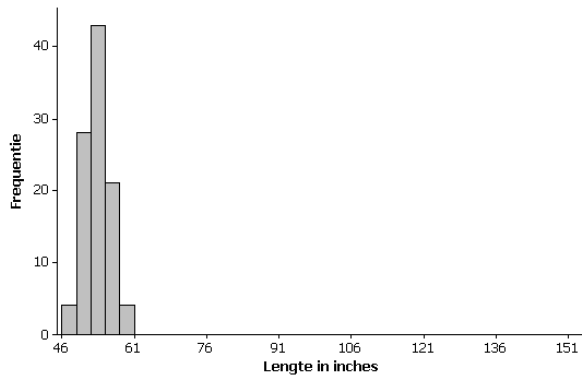
50.50 52.10 52.50 55.40 57.10 47.40 55.10 54.80 54.40 53.10 52.80 52.00
 49.40 54.80 50.60 52.00 55.20 51.60 56.30 50.50 50.90 50.00 52.40 55.30
 54.30 52.10 49.10 55.70 56.10 51.20 52.40 54.00 53.40 50.60 46.20 52.40
 54.00 50.70 50.90 49.80 54.00 48.70 50.50 50.40 54.70 52.10 50.20 56.40
 56.00 52.30 54.40 48.70 56.00 51.30 53.00 57.50 50.60 52.90 50.70 54.00
 54.10 56.10 56.80 57.00 54.00 53.60 53.50 51.00 52.30 56.10 54.90 49.30
 52.00 58.00 54.50 52.20 51.70 59.40 50.70 54.80 51.40 58.00 54.30 53.50
 52.80 57.10 53.20 50.40 51.30 51.30 56.40 51.10 52.90 52.80 58.20 56.00
 55.00 53.60 55.70 52.40



De vorm van het histogram leert je dat het gemiddelde \bar{x} als een “typische” lengte van die groep kinderen kan beschouwd worden. Hier is $\bar{x} = 53.13$ inches.

Waarschijnlijk kan je je daar heel weinig bij voorstellen wanneer je weet dat een kind 53 inches groot is. Dat is niet verwonderlijk, want jij bent gewoon om lengten in meter uit te drukken. Een groot probleem is dat niet zodra je weet dat 1 inch = 2.54 cm. Zo kan je van de oude x_i getallen (in inches) overstappen op nieuwe y_i getallen (in cm) en het verband wordt hierbij gegeven door $y_i = 2.54 x_i$.

Wat is nu de gemiddelde lengte van die 100 kinderen, maar dan in cm? Zoals vroeger wil je liever niet die 100 lengten eerst omrekenen naar cm. Bedenk eens wat er gebeurt als je alle getallen vermenigvuldigt met 2.54. Alle x_i 's zijn positief, en vermenigvuldigen met 2.54 levert y_i 's die dus groter zijn. Maar zij liggen ook meer uiteengespreid. Het histogram voor de x_i 's loopt van 46 tot 60, en vermenigvuldigen met 2.54 leidt tot een histogram dat van 116 tot 153 loopt. Als je dat op eenzelfde figuur tekent ziet het er als volgt uit. Om je te overtuigen dat het histogram van de x_i 's hetzelfde is als hierboven, kan je dit eerst eens hertekenen op een andere schaal.



Als je het oude gemiddelde $\bar{x} = 53.13$ nu ook vermenigvuldigt met 2.54 dan vind je 134.9 en dat is blijkbaar ook het evenwichtspunt van het nieuwe histogram in stippellijn.

Wat je hier grafisch hebt ontdekt is inderdaad algemeen waar. Als je elk getal x_i met een vast getal a vermenigvuldigt (of er door deelt, met dan natuurlijk $a \neq 0$) dan krijg je nieuwe getallen y_i . Het gemiddelde \bar{y} van die nieuwe getallen vind je door het oude gemiddelde \bar{x} met dat vaste getal a te vermenigvuldigen (of er door te delen).

Voor elk vast getal a geldt:

$$y_i = ax_i \quad (\text{voor } i = 1, 2, \dots, n) \Rightarrow \bar{y} = a\bar{x}$$

Je kan deze eigenschap natuurlijk ook eenvoudig bewijzen, want je weet dat $y_i = ax_i$. Je

hebt voor de som: $\sum_{i=1}^n y_i = \sum_{i=1}^n ax_i = ax_1 + ax_2 + \dots + ax_n = a(x_1 + x_2 + \dots + x_n) = a \sum_{i=1}^n x_i$

Deel tenslotte de som door het aantal en je vindt: $\bar{y} = \frac{1}{n} a \sum_{i=1}^n x_i = a \frac{1}{n} \sum_{i=1}^n x_i = a\bar{x}$.