

## De rol van leergang-specifieke software bij begripsvorming in aanvankelijk statistiekonderwijs

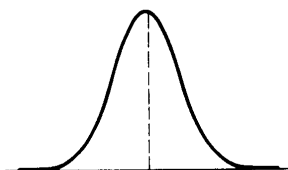
Koeno Gravemeijer

Wanneer apparaten statistische bewerkingen uitvoeren, is het van belang dat gebruikers de onderliggende statistiek goed begrijpen. Statistiek is echter een lastig onderwerp en we kunnen ons afvragen wat de leerlingen nu precies zouden moeten begrijpen en wat er op dit gebied haalbaar is. De gedachte die we in dit artikel willen uitwerken is dat computerprogramma's, die daar specifiek voor zijn ontworpen, hier een belangrijke rol kunnen spelen. We zien daarbij met name mogelijkheden in programma's die de leerlingen als gereedschap kunnen gebruiken bij het oplossen van statistische problemen en tegelijkertijd onderdeel zijn van een leergang en zo zijn ontworpen dat ze het leerproces ondersteunen. Kern hierbij is dat deze programma's startfunctionaliteiten biedt die direct aansluiten op wat de leerlingen al weten en kunnen. Door gerichte opdrachten kunnen de leerlingen, al werkend met de programma's, nieuwe inzichten ontwikkelen. Dan kunnen meer gesofisticeerde representaties en nieuwe opties worden aangeboden, die de leerlingen in de gelegenheid stellen deze inzichten verder uit te bouwen en op een hoger niveau te brengen. Wat weer mogelijkheden biedt voor nieuwe representaties etc. Door gebruik te maken van dergelijke programma's kunnen statistische inzichten eerder en effectiever worden aangeboden. Hoe specifiek op de leergang toegesneden programma's het leerproces kunnen ondersteunen lichten hieronder toe aan de ervaringen die zijn opgedaan met een experimentele leergang data analyse die gebruik maakt van speciaal daarvoor ontworpen computerprogramma's die bekend staan als de "Minitools" (Gravemeijer, 2002). Dit experiment heeft plaats gevonden in de Verenigde Staten, maar een vergelijkbare leergang was ook onderwerp van een onderwijsexperiment in Nederland (Bakker, 2004). De gebruikte Minitools zijn helaas niet meer beschikbaar, maar vergelijkbare activiteiten zijn mogelijk met op dit moment beschikbare programma's, zoals die te vinden zijn het in Digitale Wiskunde Omgeving (DWO) van het Freudenthal Instituut, of van het computerprogramma Tinkerplots (<https://www.tinkerplots.com>). Echter deze programma's zijn niet toegesneden op de hieronder beschreven leergang. Ze missen daarom specifieke kenmerken, we komen daar later op terug.

### *het concept 'verdeling' als onderwijsdoel*

De Minitools zijn ontwikkeld in het kader van een onderwijsproject in Nashville dat zich richtte op Middle-School leerlingen (vergelijkbaar met het eerste en tweede leerjaar in het VO). Het statistiekonderwijs voor die leeftijdsgroep betreft grofweg de introductie van kengetallen als gemiddelde, mediaan, modus, kwartielen en spreiding en het werken met representaties als histogram en box plot. In het gangbare onderwijs ligt het accent in het algemeen op procedures: berekenen, tekenen en aflezen. Het ontwikkelen van inzicht in betekenissen blijft daar vaak bij achter. Wanneer we op zoek gaan naar betekenissen kunnen we ons bijvoorbeeld afvragen, waar de bovengenoemde kengetallen, kengetallen van zijn? Wel, deze getallen hebben betrekking op verzamelingen van meetwaarden; data sets. De kengetallen en de visualiseringen beschrijven hoe de data/meetwaarden zijn verdeeld. Anders gezegd, het zijn hulpmiddelen om verdelingen te beschrijven. Wanneer we ons willen richten op de conceptuele kern van het aanvankelijke statistiek onderwijs, zullen we ons dus moeten richten op het begrip verdeling. Genoemde Minitools zijn ontworpen om de leerlingen te helpen dit concept, verdeling, te ontwikkelen.

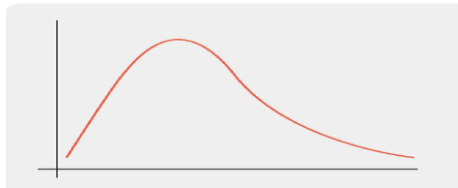
Wat we onder een verdeling verstaan kunnen we toelichten aan de normale verdeling. Iedereen kent wel het plaatje van de klokvormige kromme (figuur 1).



Figuur 1. Normale verdeling.

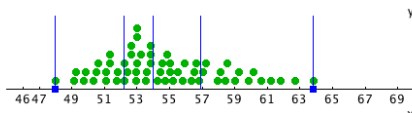
Kenmerkend aan de normale verdeling is de symmetrie en het feit dat de meeste data in het midden liggen en de aantallen afnemen als je verder van het midden (het gemiddelde) komt. Van belang is dus de vorm van de verdeling. Verder zijn uiteraard ook de meetwaarden waar het omgaat van belang; die

worden bijvoorbeeld beschreven met het gemiddelde en de spreiding t.o.v. het gemiddelde. Daarbij is de normale verdeling symmetrisch, maar verdelingen kunnen ook scheef zijn (figuur 2).

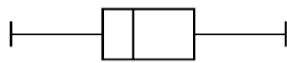


Figuur 2. Scheve verdeling.

In die zin kunnen we verdeling opvatten als een object met bepaalde kenmerken; positie, scheefheid en spreiding. We kunnen deze kenmerken beschrijven met de mediaan, kwartielwaarden en extreme waarden (zie figuur 3). Die we weer kunnen samenvatten in een box plot (figuur 4).

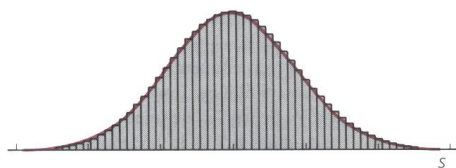


Figuur 3. dot plot met mediaan, kwartielwaarden en extreme waarden.



Figuur 4. Box plot

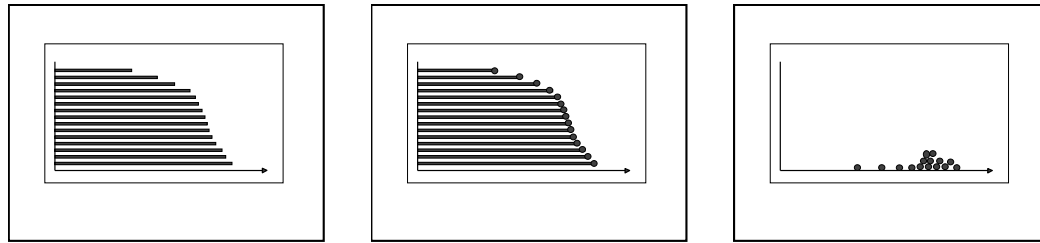
Het object 'verdeling' kunnen we zien als een kromme (een functie) die de dichtheid van datapunten beschrijft; hoe meer meetwaarden dicht bij elkaar liggen hoe hoger dat deel van de kromme (de functie waarde). Meer formeel geredeneerd beschrijft de kromme de het de limiet van het aantal datapunten per interval wanneer de intervalbreedte naar nul gaat (figuur 5).



Figuur 5. Histogram van het aantal datapunten per interval.

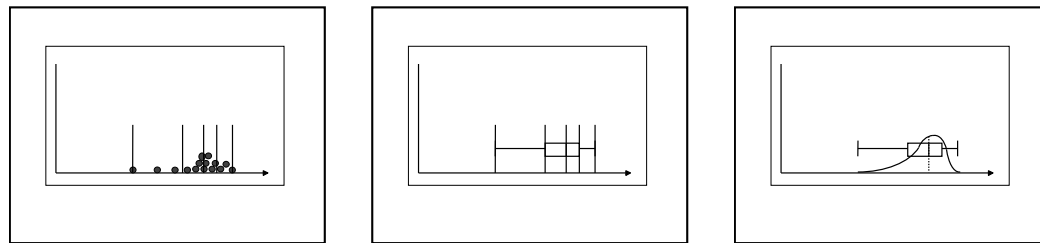
### ***grote lijn onderwijstraject***

Op basis van deze analyse kunnen we concluderen dat een meer conceptuele invulling van aanvankelijke statistiek vraagt dat meer aandacht wordt besteed aan de notie van verdeling als object. De idee achter de leergang die we hier kort beschrijven is dat de leerlingen kengetallen niet kant-en-klaar krijgen aangeboden, maar deze samen met het concept verdeling ontwikkelen. Daarbij wordt de volgende opbouw gevolgd. Eerst worden meetwaarden voorgesteld met horizontale staven waarvan de lengte evenredig is met de meetwaarde (een soort gekantelde staafgrafiek). Wanneer leerlingen zo gerepresenteerde data sets vergelijken, gaan ze al snel zien dat het gaat om de positie van de eindpunten van de staven ten opzichte van de x-as. In de volgende representatie worden de staven daarom weggelaten en wordt volstaan met de verzameling eindpunten (figuur 6).



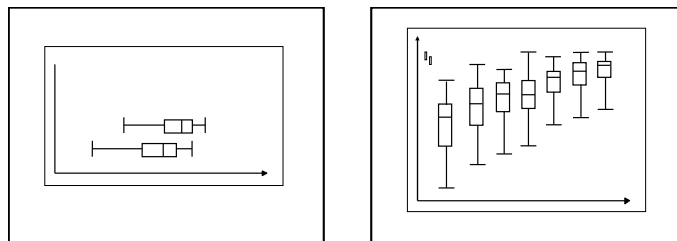
Figuur 6. Van staven (value bars) naar punten (dot plot).

Deze dot plot kunnen met behulp van de computertools op verschillende gestructureerd worden. Bijvoorbeeld door de datapunten in te delen in vier gelijke groepen. Welke weer kan worden vervangen door een box plot, waarin de leerlingen idealiter de vorm van de verdeling nog kunnen zien (figuur 7).



Figuur 7. Indelen in vier gelijke groepen als verbinding tussen de vorm van de verdeling en de box plot.

Met behulp van deze representaties kunnen twee of meer (uni-variate) verdelingen worden vergeleken, waarmee een begin kan worden gemaakt van het onderzoeken van data die afhangen van twee variabelen (bi-variate verdelingen) (Figuur 8).

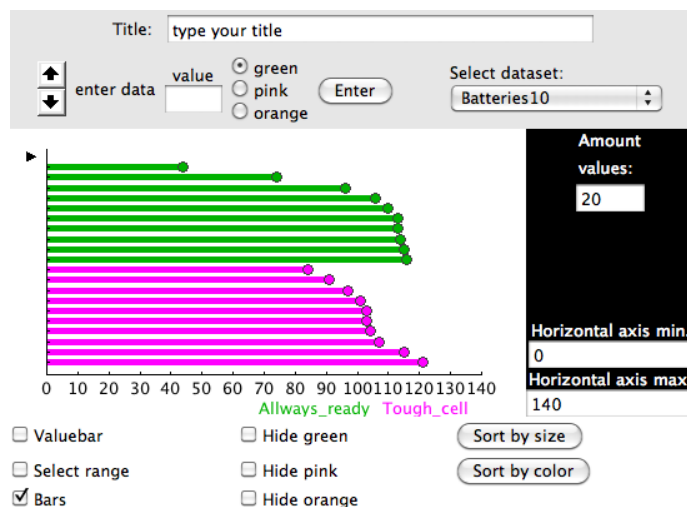


Figuur 8. Van uni-variate verdelingen naar een verkenning van bi-variate verdelingen.

**evaringen met de Minitools**

**1. meetwaarden als stroken**

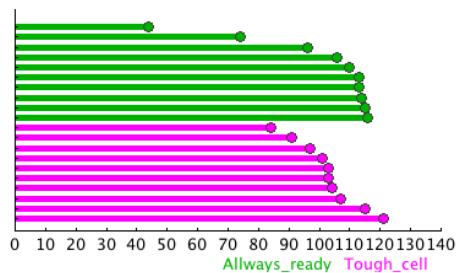
Deze representaties zijn onderdeel van de eerder genoemde Minitools, die de leerlingen opties bieden om de data sets te structureren (Figuur 9).



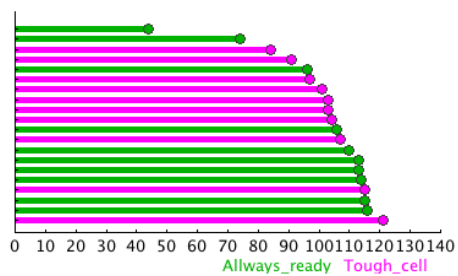
Figuur 9. Levensduur van verschillende merken batterijen.

Een *valuebar* en een *range tool* in het geval van minitool 1, waarmee de leerlingen een specifieke waarde in de grafiek kunnen markeren, respectievelijk de grenzen van een interval kunnen markeren. Het idee was dat de leerlingen de value bar zouden gebruiken om visueel het gemiddelde te bepalen en de range tool zouden gebruiken om de modus te bepalen. Dit deden ze niet, maar deze hulpmiddelen vervulden wel een nuttige functie, zoals uit het volgende voorbeeld blijkt.

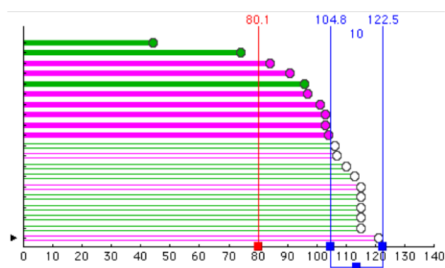
De Middle School leerlingen uit Nashville werd gevraagd om een commentaar te bedenken voor een consumentenrapport dat handelde over verschillende merken batterijen. In dit geval ging het om batterijen van het merk Always Ready en batterijen van het merk Tough Cell. Metingen van de levensduur van 10 batterijen van beide merken waren uitgezet in Minitool 1. daarbij konden de leerlingen de data splitsen in twee groepen (zie figuur 10) of doorelkaar in volgorde plaatsen (figuur 11).



Figuur 10. Minitool 1, data in twee groepen gesplitst.



Figuur 11. Minitool 1, data geordend naar grootte.



Figuur 13. Minitool 1, tien staven met hoogste waarde geselecteerd, value bar op 80.1.

In een klassengesprek gebruikten ze de tweede representatie. Een van de leerlingen gebruikt de range tool om de tien staven met de hoogste waarde te selecteren (Figuur 10). Verwijzend naar het resulterende beeld, hield ze de volgende redenering.

*And I was saying, see like there's seven green that last longer.*

Een medeleerling licht toe:

*She's saying that out of ten of the batteries that lasted the longest, seven of them are green, and that's the most number, so the Always Ready batteries are better because more of those batteries lasted longer.*

Een andere leerling was het hier niet mee eens en plaatste de value bar op 80. Hij redeneerde als volgt:

*See, there's still green ones [Always Ready] behind 80, but all of the Tough Cell is above 80. I would rather have a consistent battery that I know will get me over 80 hours than one that you just try to guess.*

Een derde leerling ging hierop in door op te merken, dat het van belang was te weten waar je de batterijen voor gebruikt.

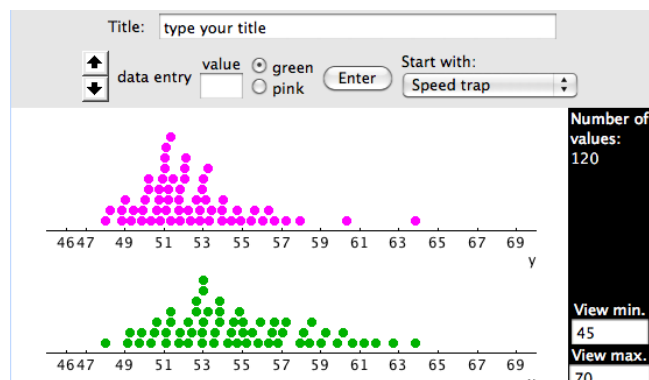
*Like, if you're using them for something real important and you're only going to have like one or two batteries, then I think you need to go with the most constant thing. But if you're going like, "Oh well, I just have a lot of batteries here to use," then you need to have most of the highest.*

Interessant is hier dat de leerlingen in samenhang met het oplossen van dit type problemen een informele spreidingsmaat uitvonden, die ze aanduidden met het woord, “consistent”.

## 2. dichtheid en vorm

Bij Minitool 2 konden de leerlingen de data structureren door verticale markeerlijntjes aan te brengen. Dit kon op verschillende manieren (a) zelf naar eigen inzicht aanbrengen, (b) de computer de data set te laten structureren, door deze in twee, of vier, gelijke delen te laten splitsen, of (c) aan door de computer op te geven hoe groot de intervallen moesten worden. De tweede manier werd onder meer gebruikt bij de volgende opgave (figuur 14).

Onderstaande dot plots tonen de gemeten snelheden van auto's voor en na de campagne, “Matig uw snelheid”. De verticale staven delen de datapunten in vier gelijke groepen.  
> Wat kun je zeggen over het effect van deze campagne?

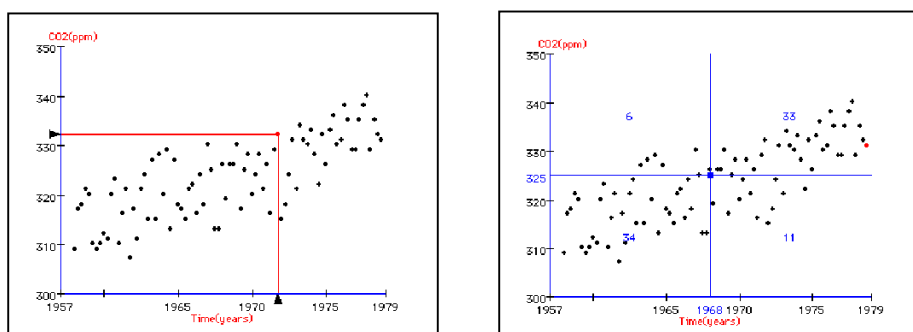


Figuur 14. Gemeten snelheden voor en na een veilig-verkeer actie.

De leerlingen merkten op dat na de campagne nog maar  $\frac{1}{4}$  van de auto's sneller dan 54 mijl/uur ging, terwijl het daarvoor de helft was. een andere beschrijving kwam van een leerling die opmerkte dat, “*the hill has shifted*”. Ze legde uit dat top van de heuvel naar voren was verschoven en dat dit betekende dat de auto's langzamer waren gaan rijden. Belangrijk is hier om ons te realiseren dat de vorm van de verdeling betekenis voor haar had. Een betekenis die de leerlingen zelf hadden moeten construeren, omdat er geen verticale as gegeven was. De vorm van de verdeling krijgt via dit soort opdrachten geleidelijk aan het karakter van een *dichtheidsfunctie*. Wanneer met een unimodale verdeling wordt gewerkt is de mediaan een goede indicator is van de *top of the hill* en de mate waarin twee staafjes dicht op elkaar zitten een maat zijn voor in hoeverre “*the data are buched up*”, ofwel voor de dichtheid. Het idee is dat wanneer vier gelijke groepen representatie van figuur 10 wordt vervangen door een box plot, de leerlingen de vorm van de verdeling in de box plot blijven zien. De zo opgedane inzichten kunnen vervolgens worden gebruikt om de samenhang tussen twee variabelen te onderzoeken.

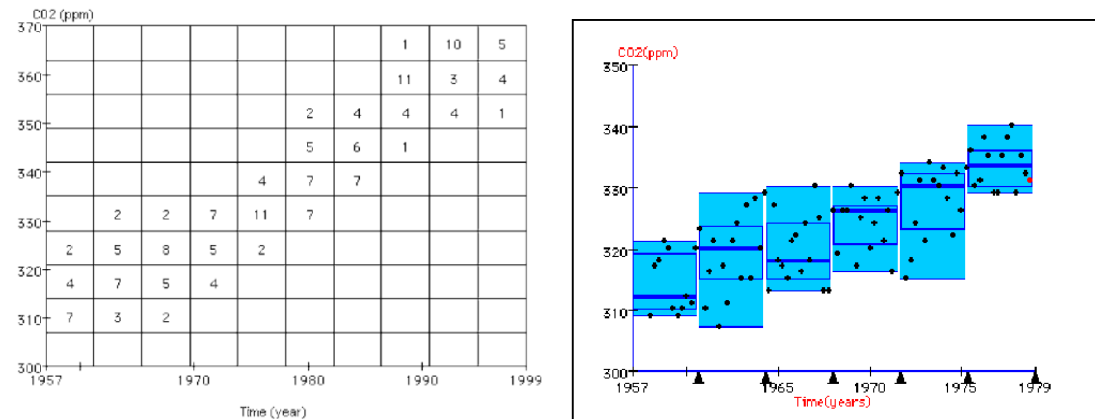
## 3. bi-variate data

Dit gebeurt in het vervolg van de leergang waarbij data die twee variabelen bevatten worden onderzocht met behulp van Minitool 3. Dan worden bijvoorbeeld gekeken naar het CO<sub>2</sub>-gehalte in opeenvolgende jaren. In Minitool 3 worden de data geplot in een vlak, in dit geval met op de verticale as het CO<sub>2</sub>-gehalte en op de horizontale as de jaren (figuur 15a). Een manier om de data te structureren is dan door middel van vier segmenten (figuur 15b).



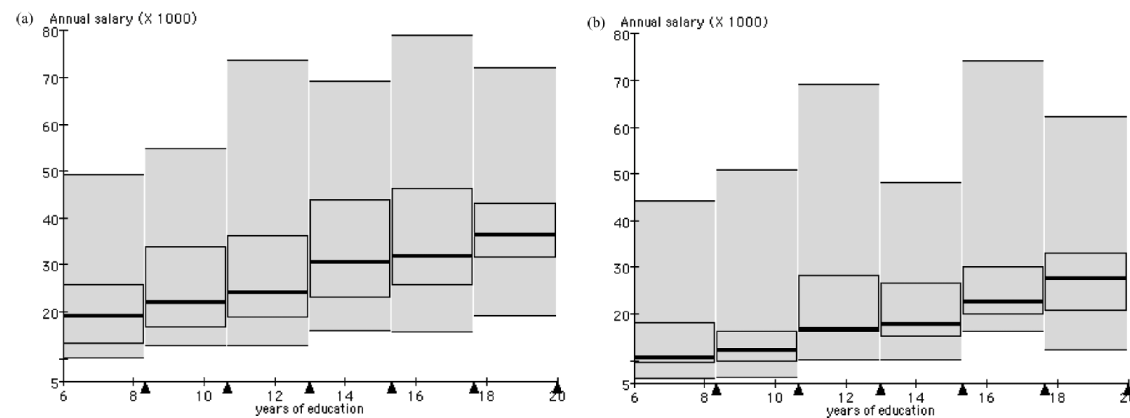
Figuur 15 a & b. CO<sub>2</sub>-gehalte uitgezet tegen kalenderjaren.

Een andere manier, door middel van een grid met aantallen data punten (figuur 16a) en een vierde door middel van verticale box plots (figuur 16b).



Figuur 16 a& b. Structureringen van CO<sub>2</sub>-gehalte uitgezet tegen kalenderjaren.

Zo'n box plot representatie kan ook worden gebruikt om twee data sets te vergelijken. Bijvoorbeeld wanneer we trends de salarisontwikkeling van vrouwelijke en manlijke leraren in een bepaald schooldistrict willen beschrijven.



Figuur 17 a & b. Salarisontwikkeling van leraren. (Gefingeerde data.)

### Reflectie

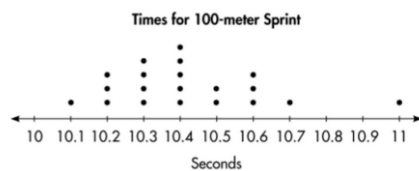
Het voorbeeld van de Minitools laat zien hoe passende software kan worden ingezet om wiskunde, of in dit geval, statistiek, te gaan begrijpen. Dit laatste achten we van belang in de 21<sup>e</sup> eeuw. Wanneer apparaten het wiskundewerk doen, dan wel allerlei statistische bewerkingen uitvoeren, dan moeten de gebruikers goed begrijpen wat de computer doet. Dit geldt enerzijds als je zelf bepaalde software gebruikt, maar ook als je de resultaten moet beoordelen van berekeningen die een ander door een computer heeft laten uitvoeren. Met bovenstaande leergang willen we laten zien dat specifiek voor dit doel ontworpen computersoftware kan helpen om dit inzicht te ontwikkelen. Daarbij gaat het om de kerninzichten die de basis voor de bewerkingen vormen. Precies bij dit type inzicht kan dynamische, interactieve software een belangrijke rol spelen om deze inzichten op een efficiënte wijze te ontwikkelen.

Uitgangspunt is hier dat het accent in het onderwijs niet moet liggen op het berekenen van kengetallen als gemiddelde, mediaan, modus, kwartielen en spreiding, of het tekenen en aflezen van statistische representaties als histogram en box plot. In plaats daarvan zullen we ons moeten richten op het begrip verdeling. Waarbij de verdeling wordt gezien als een object met bepaalde kenmerken. De Minitools zijn

ontworpen om de leerlingen te helpen het concept verdeling en de daarbij passende kengetallen en representaties op een inzichtelijke manier te ontwikkelen.

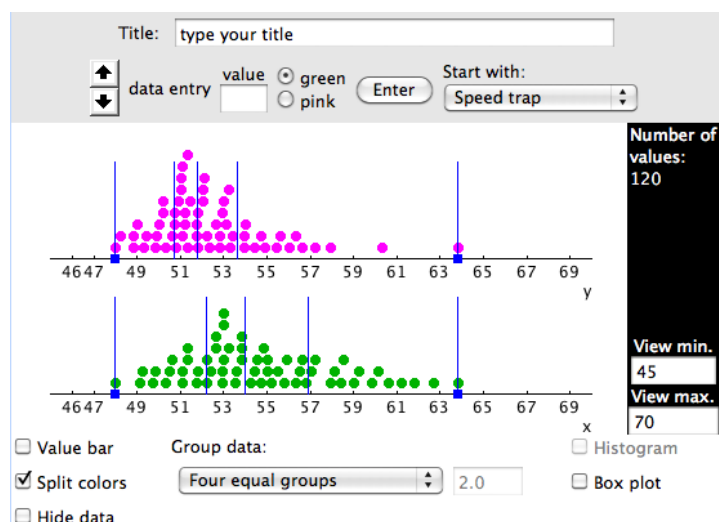
De leerlingen ervarende stroken in Minitool 1 als vanzelfsprekende visualiseringen van meetwaarden, ze zijn stroken namelijk eerder in een dergelijke rol tegen gekomen; bij de schaal van een kaart. De opties van de Minitool, als de *valuebar* en de *range tool* kunnen de leerlingen gebruiken om structuur aan te brengen en om over de data sets te redeneren. Met name door het gebruik van deze opties leerlingen ontdekken dan dat het gaat om de posities van de eindpunten van de staven. Wat vervolgens een meer compacte representatie mogelijk maakt. Bij de dot plot van Minitool 2 richten de opdrachten zich op de manier waarop deze punten (de data) verdeeld zijn; hoe de dichtheid van de datapunten varieert over de x-as. Met name de optie, de data in te delen in vier gelijke groepen richt de aandacht op de dichtheid; hoe dichter de verticale begrenzingen bij elkaar liggen, des te hoger de dichtheid. Aldoende wordt de basis gelegd voor de notie van verdeling als dichtheidsfunctie. Door de opties van Minitool 2 te gebruiken, kunnen de leerlingen statistische maten ontwikkelen als mediaan, kwartielen, en extreme waarden, en representaties ontwikkelen als box plot (of het hier niet besproken histogram) als manieren om kenmerken van verdelingen te beschrijven. De verdeling krijgt dan het karakter van een object met specifieke kenmerken, zoals symmetrie of scheefheid, spreiding en positie op de x-as. Daarmee wordt tevens de basis gelegd voor het redeneren met verdelingen in de context van covariantie. De Minitools helpen de leerlingen zo op een efficiënte manier met het begrijpen van de kernideeën achter het gebruik van statistische gereedschappen als mediaan, modus, kwartielen, spreiding, histogram, box plot en correlatie.

Zoals eerder opgemerkt verschillen de Minitools op bepaalde punten van de gangbare statistiek programma's. Zo zijn de eindpunten van de staven van Minitool 1 voorzien van bolletjes. Hiermee wordt geanticipeerd op de overgang naar Minitool 2. Deze bolletjes kunnen als het ware neerdalen op de x-as. Daarom ziet de dot plot van Minitool 2 er ook anders uit dan de gangbare dot plots, waar de bolletjes in rechte stapeltjes zijn ondergebracht (zie figuur 18). In de wat rommelige dot plots van Minitool 2 hebben de bolletjes hun exacte positie ten opzichte van de x-as behouden.



Figuur 18. Gangbare vorm van een dot plot.

Uniek zijn ook de range tool en de value bar van minitool 1. Met de range tool kunnen de leerlingen een groep datapunten markeren en we zien hoe de leerlingen hier gebruik van maken om hun redeneringen te ondersteunen. Redeneringen die leiden tot het idee van "consistentie", dat gezien kan worden als een informele versie van het concept spreiding. De value bar wordt op een vergelijkbare manier gebruikt. Een specifiek kenmerk van Minitool 2 is dat de leerlingen de data kunnen structureren door verticale markeerlijntjes aan te brengen. Dit kan op verschillende manieren. De eerste manier, zelf naar eigen inzicht aanbrengen, sluit aan bij werkwijzen waarvan bekend is dat sommige leerlingen die spontaan kiezen als ze een structurering op papier willen aanbrengen. Van belang is hier dat de leerlingen de voor en nadelen van verschillende manieren van structureren tegen elkaar kunnen afwegen. De optie om de computer de data set in twee, of vier, gelijke delen te laten splitsen, biedt de mogelijkheid om relatieve vergelijkingen te maken. Door gebruikte maken van de vier-gelijke-groepen optie bij de data over de snelheids campagne bijvoorbeeld, ontdekken de leerlingen dat eerst de helft van de auto's harder dan 54 km/u ging en later minder dan de helft (zie figuur 19).



Doordat leerlingen verschillende opties kunnen kiezen bij het oplossen van dezelfde opgave, kunnen er klassendiscussies ontstaan, waarin dezelfde verdelingen op verschillende manieren met elkaar worden vergeleken. Op deze manier kunnen ideeën over dichtheid, vorm en indeling met elkaar worden verbonden. Zo kunnen, bijvoorbeeld, noties over vorm gekoppeld worden aan dichtheid, mediaan en kwartielen.

Een kernprincipe in de leergang is dat nieuwe symbolische representaties niet uit de lucht komen vallen, maar een geschiedenis hebben voor de leerlingen. Zo is het idee, bijvoorbeeld, dat de leerlingen de bolletjes in de dot plot van Minitool 2 zien als representanten van staafjes waar deze bolletjes het eindpunt van vormden. De positie van het bolletje vertegenwoordigt daarmee de oorspronkelijke meetwaarde. Op eenzelfde manier ontleen de vier-groepen-indeling en de box plot hun betekenis aan ervaringen met het structureren van verzamelingen van datapunten en het geleidelijk opkomende idee van verdeling als object.

Tot slot willen we nog opmerken dat de opgaven voorbeelden bieden van het type toepassingsituaties waar we de leerlingen op willen voorbereiden. Bij de batterijen bijvoorbeeld gaat het erom te bedenken dat de bruikbaarheid van het gemiddelde afhangt van de vraag waar je die batterijen voor gebruikt en dat soms de spreiding een betere maat is. Bij de snelheden gaat het om het inzicht dat een box plot een heel efficiënte weergave van de data kan zijn. Mits je goed begrijpt waar een box plot voor staat. De CO<sub>2</sub> data en de leraar-salarissen laten tenslotte zien dat bij covariantie niet alleen het verband zelf belangrijk is, maar ook hoe de data rond de centrale trend zijn verdeeld.

Afsluitend kunnen we concluderen dat wanneer we het statistiekonderwijs zo willen inrichten dat het de leerlingen goed voorbereid op de maatschappij van de 21<sup>e</sup> eeuw, tailormade software daar een belangrijke rol bij kan spelen. Dit pleit voor investeren in het ontwikkelen van dit type software en vraagt ons bij het doordenken van een curriculum voor de toekomst rekening te houden met de mogelijkheden die software biedt.

### ***literatuur***

- Gravemeijer, K. (2002). Developmental research, a course in elementary data analysis as an example. In: F. Lin (Ed.) *Common Sense in Mathematics Education*, (p.p. 43-68). Taipei, Taiwan: National Taiwan Normal University.
- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht: CD-β Press.