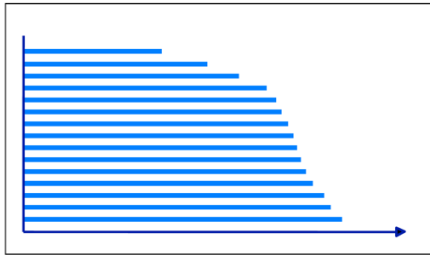


Emergent modelleren, Data-analyse als voorbeeld

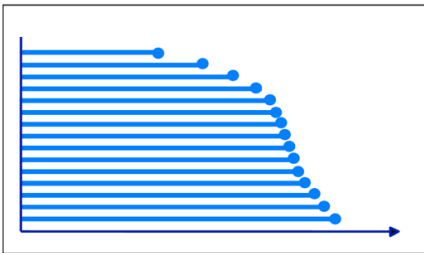
In een data-analyse leergang (2002), staat het *proces* van het construeren en analyseren van representaties van datasets centraal. Deze start met representaties van individuele meetwaarden en eindigt met de grafiek van een dichtheidsfunctie. De representaties ontwikkelen zich van model van meetwaarden naar een model voor de verdeling als wiskundig *object*.

In het eerste sub-model worden meetwaarden gerepresenteerd als horizontale lijnstukken (figuur #).

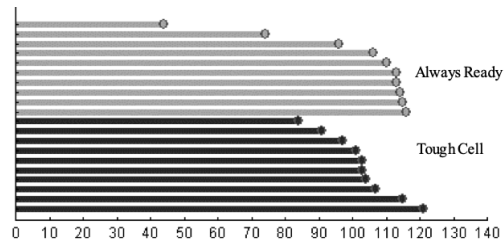


Figuur #. Met lijnstukken voorstellen van meetwaarden.

Bij het analyseren en vergelijken van datasets kijk je vooral naar de positie van de eindpunten van deze lijnstukken (Figuur # a). Bijvoorbeeld wanneer de resultaten van metingen aan twee sets batterijen worden vergeleken met het oog op een consumentenadvies (figuur # b).

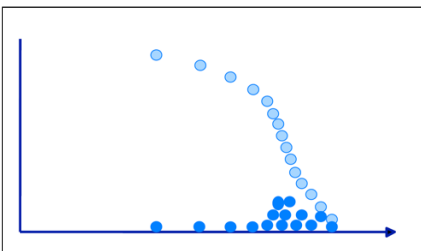


Figuur # a. De eindpunten als referenties voor de meetwaarden.



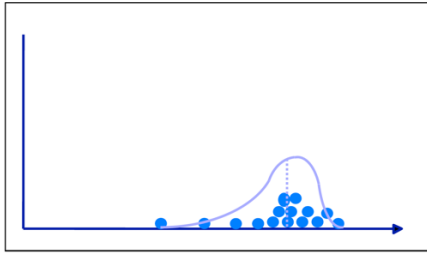
Figuur # b. Levensduur van batterijen.

Hierna kan een nieuw sub-model worden geïntroduceerd, waarin de lijnstukken worden weggelaten en alleen de eindpunten overblijven, en die naar beneden zakken, waardoor een dot plot ontstaat.



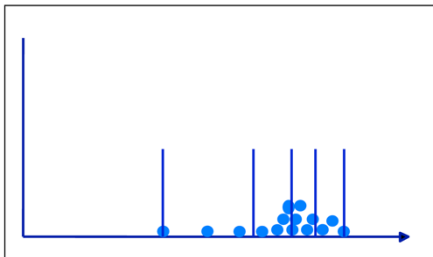
Figuur #. Van eindpunten naar dot plot.

Waarbij de 'dots' voor de leerlingen verwijzen naar de eindpunten van de lijnstukken die op hun beurt weer de meetwaarden symboliseren. Daarmee wordt de basis gelegd voor het redeneren over de vorm van de verdeling

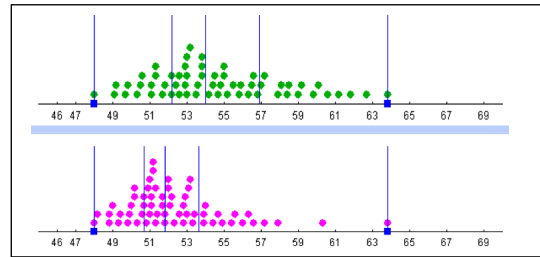


Figuur #. De vorm van de verdeling.

De dot plot representatie wordt verder gestructureerd door de datapunten in vier gelijke groepen te verdelen (figuur # a). Zoals, bijvoorbeeld, in de opgave van figuur # b waar snelheden voor en na een veilig-verkeer actie worden vergeleken.

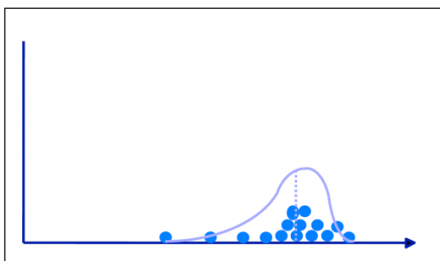


Figuur # a. Vier gelijke groepen.



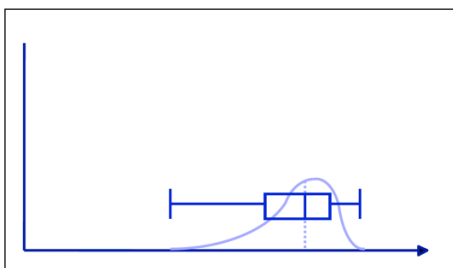
Figuur # b. Snelheden voor en na.

Eén van de leerlingen beschrijft de verandering als, ‘The hill has shifted’. Zij interpreteert de hoogte hier als een maat voor de dichtheid van de gemeten snelheden en concludeert dat men langzamer is gaan rijden. Met dergelijke observaties verschuift de aandacht naar de vorm van de verdeling en naar de verdeling als object (Figuur #).



Figuur #. Verdeling als object.

Waarbij de box plot als het meest abstracte sub-model kan worden geïntroduceerd als een manier om een verdeling te beschrijven. En waar je in het geval van een uni-variate verdeling de vorm van de verdeling kunt aflezen (figuur #).



Figuur #. Box plot en verdeling

De verdeling als object vertegenwoordigt een knooppunt in een netwerk van relaties. De vorm is een belangrijk kenmerk van de verdeling, die symmetrisch of scheef kan zijn, een grote of een kleine spreiding kent en gekoppeld is aan de waarden op de x-as. Van belang is inzicht in de relaties tussen de vorm en de manier waarop de data zijn verdeeld over een mogelijke-uitkomstenruimte. (Hier één-dimensionaal, samenvallend met de x-as.) Andere relaties die hiermee samenhangen betreffen het concept dichtheid, de relatie tussen dichtheid hoogte, de relatie tussen een datapunt en de gerepresenteerde meetwaarde en de koppeling daarvan aan de positie ten opzichte het nulpunt van de x-as.

Terzijde merken we op dat het emergent modelleren niet alleen wiskundige objecten oplevert, maar vaak ook procedures of algoritmen. Naar mate een leergang vordert worden de representaties niet alleen steeds schematischer, maar wordt het oplossingsproces meestal ook steeds verder verkort en gesystematiseerd.

Gravemeijer, K. (2002, July). Emergent modeling as the basis for an instructional sequence on data analysis. In *Proceedings of the 6th International Conference on Teaching Statistics*. <https://www.fisme.science.uu.nl/publicaties/literatuur/4795.pdf>