2-2

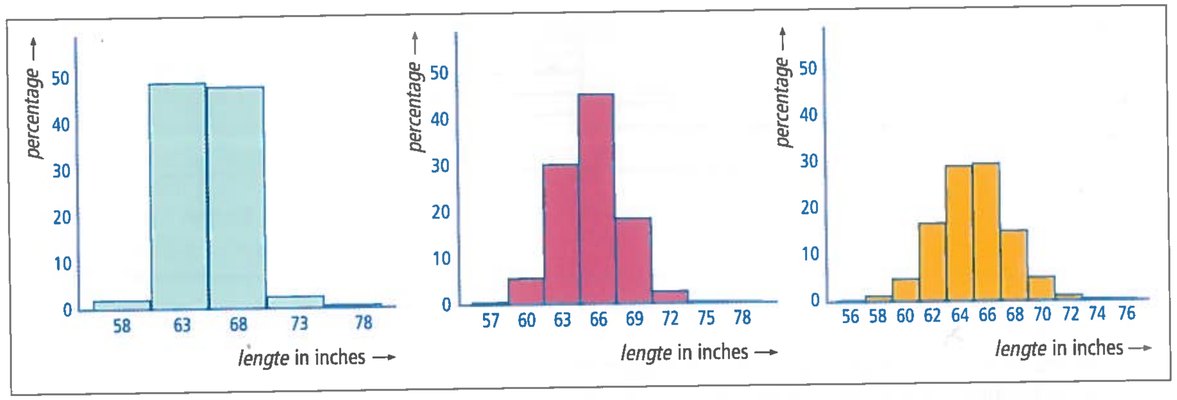
Normale verdeling

In 1901 publiceerde Scotland Yard onderstaande tabel over de lengte in inch van 3000 criminelen.

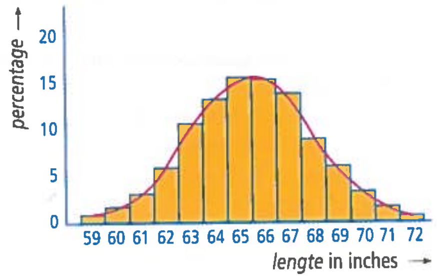
lengte in inch	56	57	58	59	60	61	62	63	64	65	66
aantal criminelen	1	1	6	23	48	90	175	317	393	462	458

lengte in inch	67	68	69	70	71	72	73	74	75	76	77
aantal criminelen	413	264	177	97	46	17	7	4	0	0	1

Bij deze tabel zijn drie staafdiagrammen getekend met klassenbreedte 5 inch, 3 inch en 2 inch.



Wat kun je zeggen over de vorm van het staafdiagram naarmate de klassenbreedte kleiner wordt?
Wat kun je dan zeggen over de hoogte van de staafjes?
Hiernaast staat het staafdiagram met klassenbreedte 1 inch. Er is een vloeiende grafiek getekend die zo goed mogelijk bij het staafdiagram past.
Waarom is het waarschijnlijk dat je bij een groep van 10 criminelen geen vloeiende grafiek kunt vinden die het staafdiagram goed benadert, maar bij een groep van 3000 criminelen wel?



De gemiddelde lengte m van de 3000 criminelen is 65,5 inch en de standaarddeviatie s is 2,5 inch.
Bereken $m - s$ en $m + s$.
Hiernaast zie je hoe Jannes het aantal criminelen tussen $m - s$ en $m + s$ berekent.
Waarom neemt hij de helft van 317 en de helft van 264?
Hoeveel procent van de criminelen heeft een lengte die tussen $m - s$ en $m + s$ ligt?

Jannes
$0{,}5 \times 317 + 393 + 462 +$
$458 + 413 + 0{,}5 \times 264 \approx \ldots$

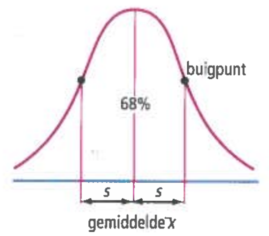
Theorie

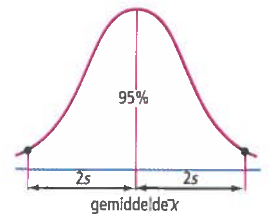
De frequentieverdeling van een variabele kun je vaak benaderen met een klokvormige grafiek.
De belangrijkste kenmerken van zo'n grafiek zijn:
> de grafiek is symmetrisch
> de symmetrie-as ligt precies bij het gemiddelde
> gemiddelde en mediaan vallen samen
> De standaardafwijking s is de afstand van de symmetrie-as tot de buigpunten van de kromme.

Er gelden de volgende drie vuistregels:
1 van de data ligt ongeveer 68% tussen $\bar{x} - s$ en $\bar{x} + s$
2 van de data ligt ongeveer 95% tussen $\bar{x} - 2s$ en $\bar{x} + 2s$
3 van de data ligt vrijwel 100% tussen $\bar{x} - 3s$ en $\bar{x} + 3s$

De ideale klokvormige verdeling wordt de **normale verdeling** genoemd. Een variabele waarvan de frequentieverdeling benaderd kan worden door de normale verdeling heet **normaal verdeeld**.

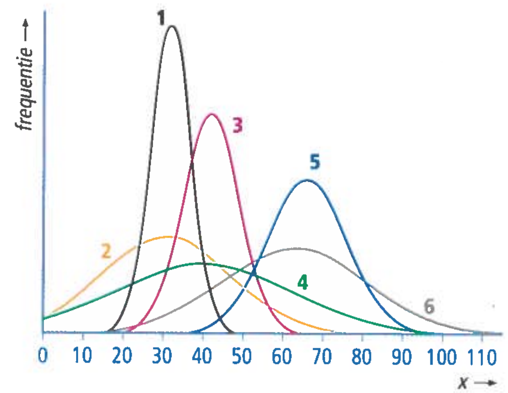
In het buigpunt verandert een toenemende stijging (of daling) in een afnemende stijging (of daling).





8 Welke van de volgende data zullen geen klokvormige verdeling opleveren? Maak in die gevallen een schets van de grafiek en licht je antwoorden toe.
A De inkomens van de Nederlandse werknemers.
B De gewichten van muntstukken van twee euro.
C De inhoud van literpakken melk.
D De lengte van alle zesjarigen en hun moeders.
E De leeftijden van alle Nederlanders.

9 In de figuur hiernaast staan zes grafieken behorende bij zes statistische variabelen die normaal verdeeld zijn.
a Welke verdeling heeft het grootste gemiddelde? En welke het kleinste?
b Welke verdeling heeft de grootste standaarddeviatie? En welke de kleinste?



De grafiek van een normale verdeling is volledig bepaald als gemiddelde en standaardafwijking bekend zijn.

10 Wat gebeurt er met de grafiek van een normale verdeling als het gemiddelde kleiner wordt en de standaarddeviatie gelijk blijft?

11 Wat gebeurt er met de grafiek van een normale verdeling als de standaarddeviatie groter wordt en het gemiddelde gelijk blijft? Verklaar je antwoord.