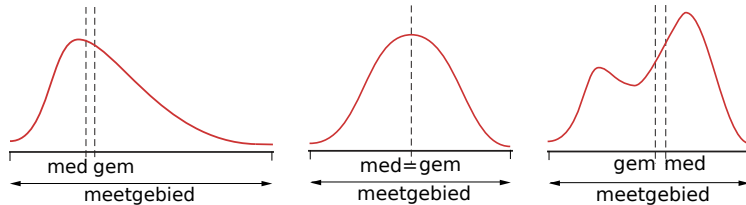


Spiekbrieftjes bij Normale verdeling

Soorten verdelingen

Bij de **vorm van een frequentieverdeling** let je op:

- de symmetrie, de scheefheid, of de gelijkmatigheid
- het aantal toppen
- de uitschieters (meer dan $1,5 \times IQR$ vanaf mediaan, of Dixon's Q-test)



Belangrijk is wel dat je bij een frequentieverdeling met een klassenindeling de mediaan en het gemiddelde alleen nog kunt schatten met behulp van de klassenmiddens. (Tenzij je de werkelijke gegevens nog hebt!)



De normale verdeling

Je ziet een **normale verdeling** van x . De **normaalkromme** wordt bepaald door het **gemiddelde** μ en de **standaardafwijking** of **standaarddeviatie** σ .

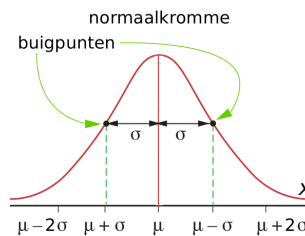
Bij een steekproef geldt het steekproefgemiddelde $\bar{x}$ en de steekproefstandaardafwijking s_x .

Bij de waarden x_i en frequenties f_i is de standaardafwijking:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu_x)^2 \cdot f_i}{n}} \quad (\text{populatie}) \quad \text{of} \quad s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{n-1}} \quad (\text{steekproef}).$$

Voor elke normale verdeling gelden de vuistregels:

- $\approx 68\%$ van alle waarden liggen tussen $\mu - \sigma$ en $\mu + \sigma$
- $\approx 95\%$ van alle waarden liggen tussen $\mu - 2\sigma$ en $\mu + 2\sigma$
- $\approx 100\%$ van alle waarden liggen tussen $\mu - 3\sigma$ en $\mu + 3\sigma$



Rekenen met normale verdelingen

Veel statistische variabelen zijn **normaal verdeeld**.

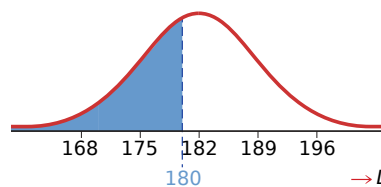
De gekleurde oppervlakte in de figuur stelt het deel voor met een lengte vanaf 165 tot 180 cm. Dat noteer je als: $P(165 \leq L < 180)$.

Omdat $P(L = 165) = 0$ is dit hetzelfde als $P(165 < L < 180)$.

De relatieve frequentie dat iemand afgerond een lengte heeft van 165 centimeter, is $P(164,5 < X < 165,5)$. Houdt bij een normale verdeling geen rekening met afrondingen, tenzij dat duidelijk uit de vraagstelling blijkt.

Alle normale verdelingen zijn om te zetten naar de **standaardnormale verdeling** met z -waarden: $z = \frac{x - \mu_x}{\sigma_x}$.

Met de **standaardnormale tabel** kun je de bijbehorende relatieve frequenties bepalen.



Betrouwbaarheidsintervallen

Het **populatiegemiddelde** wordt vaak geschat door het **steekproefgemiddelde**. Daardoor ontstaat een toevalsfout. (Met de toevalsfout in de standaardafwijking houdt je geen rekening.)

De **steekproevenverdeling** van de gemiddelden $\bar{X}$ is altijd normaal verdeeld.

De standaardafwijking S is: $S_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ met n de grootte van de steekproeven en σ de populatiestandaarddeviatie.

Het populatiegemiddelde ligt in het **betrouwbaarheidsinterval**, afhankelijk van de **betrouwbaarheid**.

- 95%-betrouwbaarheidsinterval: ondergrens = $\bar{X} - 1,96 \cdot \frac{\sigma}{\sqrt{n}}$ en bovengrens = $\bar{X} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}$.
- 99%-betrouwbaarheidsinterval: ondergrens = $\bar{X} - 2,575 \cdot \frac{\sigma}{\sqrt{n}}$ en bovengrens = $\bar{X} + 2,575 \cdot \frac{\sigma}{\sqrt{n}}$.

Conclusie: Met een betrouwbaarheid van 95 of 99% ligt het gemiddelde van de populatie tussen de 'ondergrens' en de 'bovengrens'.



meer info

De Shewhartkaart

Bij het bewaken van de kwaliteit van bepaalde processen worden **controlekaarten**, **regelkaarten** of **Shewhartkaarten** genoemd. Hier zie je er één.

- De middelste (groene) lijn bepaalt het gemiddelde μ .
- **waarschuwingsgrenzen:** $\mu \pm 2\sigma$ -lijnen met ertussen 95% van de metingen.
- **actiegrenzen:** $\mu \pm 3\sigma$ -lijnen met ertussen vrijwel alle metingen.
- **tolerantiegrenzen:** twee grenzen die absoluut niet mogen worden overschreden.

Bij dergelijke controlekaarten gelden de **Westgardregels**.



meer info