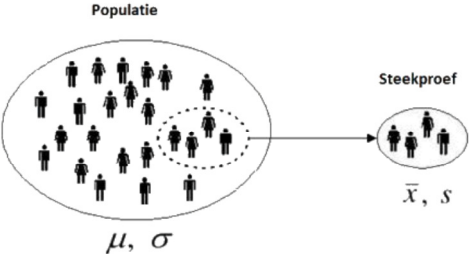


Statistische variabelen



 **HAVO wiskunde A hoofdstuk 10**
 **formuleblad**

0. voorkennis

Soorten variabelen

Bij **kwantitatieve gegevens** gaat het om meetbare gegeven, zoals temperatuur, snelheid of gewicht.

Een **kwantitatieve variabele** wordt uitgedrukt in een getal. Het verschil tussen twee van deze opeenvolgende getallen heeft een eenduidige betekenis.

Nominaal, ordinaal, interval en rationiveau

Bij de statistiek kom je vaak de volgende indeling tegen:

✓ Nominaal

Hier heeft de waarde die een variabele kan aannemen alleen de betekenis van een naam. Er is geen sprake van een volgorde. Denk bijvoorbeeld aan rugnummers van een voetbal elftal. Je kunt niet zeggen dat de speler met rugnummer 14 beter is dan de speler met rugnummer 7.

✓ Ordinaal

Een waarde op ordinaal niveau geeft alleen een volgorde aan. Denk bijvoorbeeld aan opleiding. Voorbeeld

- ✓ =MBO
- ✓ =HBO
- ✓ =UNIVERSITEIT

Er is wel sprake van volgorde, een hoger nummer duidt op een hogere opleiding, maar het verschil tussen bijvoorbeeld 1 en 2 is niet hetzelfde als het verschil tussen 2 en 3.

✓ Interval

Bij variabelen op intervalniveau hebben verschillen wel een betekenis. Neem bijvoorbeeld temperatuur. Het verschil tussen 30A en 40A is hetzelfde als het verschil tussen 70A en 80A. Je kunt echter niet zeggen dat 80A twee keer zo warm is als 40A.

✓ Ratio

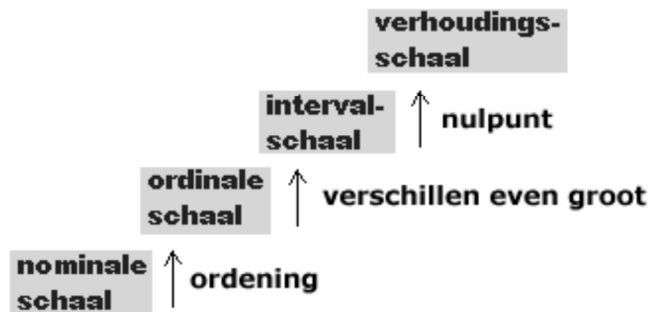
Bij variabelen op rationiveau heb je altijd een nulpunt. Denk aan gewicht, lengte of het aantal verkochte exemplaren van een product. Je kunt uitspraken doen 'A scoort twee keer zo hoog als B'.

Discreet of continu

Bij een **discrete variabele** worden alleen maar 'losse' waarden aangenomen. Tussenwaarden hebben geen betekenis.

Bij **continue variabelen** kunnen alle waarden op een interval worden aangenomen.

Nominaal, ordinaal, interval en rationiveau

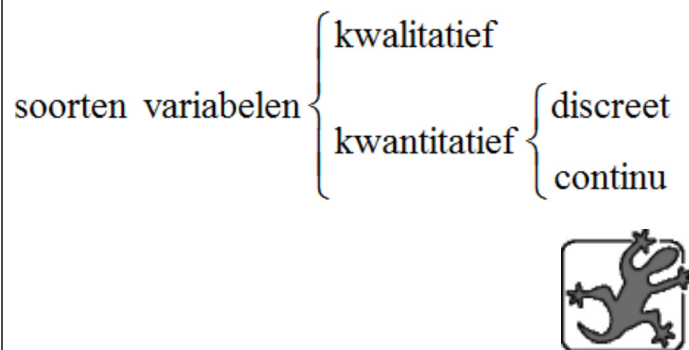


Zoals je ziet, komt er bij elk niveau hoger een eigenschap bij

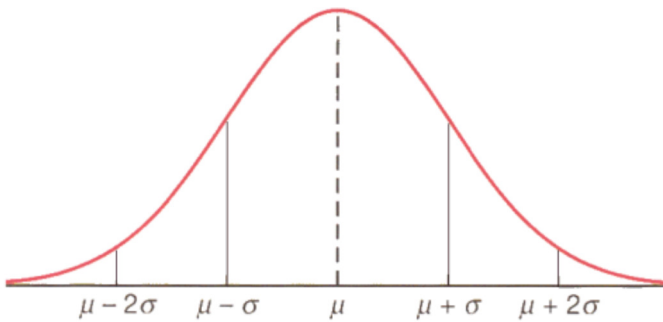
Opmerking uit het boek

"Men maakt bij kwantitatieve variabelen onderscheid tussen variabelen op intervalniveau en op rationiveau, maar deze begrippen hoef je niet te kennen op het centraal examen."

getal & ruimte - havo wiskunde A deel 3 - pagina 54



De normale verdeling



figuur 7.21

Een paar eigenschappen van een normale verdeling:

- ✓ **klokvormig**
- ✓ **symmetrisch** t.o.v. het gemiddelde.
- ✓ gemiddelde, mediaan en modus vallen samen
- ✓ de verdeling wordt bepaald door de **verwachtingswaarde** en de **standaarddeviatie**

Vuistregels bij de normale verdeling

- I. **68%** van de gegevens wijkt op z'n hoogst ± 1 keer de standaarddeviatie af van de verwachtingswaarde
- II. **95%** van wijkt op z'n hoogst twee keer de standaarddeviatie af van de verwachtingswaarde

Voorbeeld



Het gewicht van de mandarijnen uit een grote partij is normaal verdeeld met een gemiddelde van 80 gram. Verder is bekend dat 16% van de mandarijnen minder dan 76 gram weegt.

- ✓ Bereken de standaardafwijking

Uitwerking

16% weegt minder dan 76 gram, dus 76 ligt ± 1 standaardafwijking van het gemiddelde af. De standaardafwijking is $80 - 76 = 4$ gram.

1. populatie en steekproef

Betrouwbaarheidsintervallen voor de populatieproportie

Voor het berekenen van een betrouwbaarheidsinterval voor een populatieproportie moet de steekproefproportie \hat{p} en de steekproefomvang n bekend zijn.

Het 95%-betrouwbaarheidsinterval is $[\hat{p} - 2\sigma, \hat{p} + 2\sigma]$.

Voor de standaardafwijking σ ken je de formule:

$$\sigma = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Op het formuleblad

Het 95%-betrouwbaarheidsinterval voor de populatieproportie is:

$$p \pm 2 \cdot \sqrt{\frac{p(1-p)}{n}}$$

met p de steekproefproportie en n de steekproefomvang.

Afspraak

Bij de steekproeven in dit hoofdstuk gaan we er van uit dat de steekproeven aselekt en representatief zijn. Bovendien nemen we aan dat de steekproevenverdeling normaal verdeeld is.

Voorbeeld 1

Bij een steekproef van lengte 35 is het 95%-betrouwbaarheidsinterval voor het gemiddelde $[170,2; 179,8]$.

- ✓ Bereken het steekproefgemiddelde en de steekproefstandaardafwijking.

Uitwerking

- ✓ **Uitwerking voorbeeld 1**

Betrouwbaarheidsintervallen voor populatiegemiddelde

Voor een populatiegemiddelde is:

- ✓ het 68%-betrouwbaarheidsinterval:

$$\left[\bar{X} - \frac{S}{\sqrt{n}}, \bar{X} + \frac{S}{\sqrt{n}} \right]$$

- ✓ het 95%-betrouwbaarheidsinterval:

$$\left[\bar{X} - 2 \cdot \frac{S}{\sqrt{n}}, \bar{X} + 2 \cdot \frac{S}{\sqrt{n}} \right]$$

Hierin is \bar{X} het steekproefgemiddelde, S de steekproefstandaardafwijking en n de steekproefomvang.

Het 95%-betrouwbaarheidsinterval geeft aan dat het interval met 95% zekerheid het populatiegemiddelde bevat.

Op het formuleblad

Het 95%-betrouwbaarheidsinterval van het populatiegemiddelde is:

$$\bar{X} \pm 2 \cdot \frac{S}{\sqrt{n}}$$

met \bar{X} het steekproefgemiddelde, n de steekproefomvang en S de steekproefstandaardafwijking.

Voorbeeld 2

Bij een steekproef waarbij de steekproefstandaardafwijking 10 is, is het 95%-betrouwbaarheidsinterval voor het gemiddelde $[175,185]$.

- ✓ Bereken het steekproefgemiddelde en de steekproefomvang.

Uitwerking

- ✓ **Uitwerking voorbeeld 2**

 **formuleblad**

2. groepen vergelijken

<p>Verschillen kwantificeren bij nominale variabelen</p> <p>Voor het kwantificeren van een verschil bij een nominale variabele gebruiken we een associatiemaat.</p> <ol style="list-style-type: none"> 1. Het percentageverschil (PV) 2. De odds-ratio (OR) 3. De phi-coëfficiënt (phi) 	<p>Vuistregels</p> <table border="1" data-bbox="824 310 1497 537"> <thead> <tr> <th>PV</th> <th>OR</th> <th>phi</th> <th>het verschil is</th> </tr> </thead> <tbody> <tr> <td>$PV > 30\%$</td> <td>$OR > 3$</td> <td>$\phi < -0,4$ of $\phi > 0,4$</td> <td>groot</td> </tr> <tr> <td>$15\% < PV \leq 30\%$</td> <td>$2 \leq OR \leq 3$</td> <td>$-0,4 < \phi < -0,2$ of $0,2 < \phi < 0,4$</td> <td>middelmatig</td> </tr> <tr> <td>$PV \leq 15\%$</td> <td>$OR < 2$</td> <td>$-0,2 < \phi < 0,2$</td> <td>gering</td> </tr> </tbody> </table> <p>De vuistregels worden gebruikt om te bepalen of het verschil groot, middelmatig of gering genoemd wordt.</p>	PV	OR	phi	het verschil is	$PV > 30\%$	$OR > 3$	$\phi < -0,4$ of $\phi > 0,4$	groot	$15\% < PV \leq 30\%$	$2 \leq OR \leq 3$	$-0,4 < \phi < -0,2$ of $0,2 < \phi < 0,4$	middelmatig	$PV \leq 15\%$	$OR < 2$	$-0,2 < \phi < 0,2$	gering
PV	OR	phi	het verschil is														
$PV > 30\%$	$OR > 3$	$\phi < -0,4$ of $\phi > 0,4$	groot														
$15\% < PV \leq 30\%$	$2 \leq OR \leq 3$	$-0,4 < \phi < -0,2$ of $0,2 < \phi < 0,4$	middelmatig														
$PV \leq 15\%$	$OR < 2$	$-0,2 < \phi < 0,2$	gering														
<p>Kruistabel</p> <p>In de 2x2-kruistabel zijn de aantallen met de letters a, b, c en d aangegeven.</p>	<table border="1" data-bbox="831 680 1409 898"> <thead> <tr> <th></th> <th>z</th> <th>D</th> <th></th> </tr> </thead> <tbody> <tr> <th>j</th> <td>a</td> <td>b</td> <td>a+b</td> </tr> <tr> <th>m</th> <td>c</td> <td>d</td> <td>c+d</td> </tr> <tr> <th></th> <td>a+c</td> <td>b+d</td> <td></td> </tr> </tbody> </table>		z	D		j	a	b	a+b	m	c	d	c+d		a+c	b+d	
	z	D															
j	a	b	a+b														
m	c	d	c+d														
	a+c	b+d															
<p>het percentageverschil PV</p>	$PV = \frac{a}{a+b} \times 100\% - \frac{c}{c+d} \times 100\%$																
<p>de odds-ratio (OR)</p>	$OR = \frac{\frac{a}{b}}{\frac{c}{d}} \text{ ofwel } OR = \frac{ad}{bc}$																
<p>de phi-coëfficiënt (phi)</p>	$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$																
<p>Verschillen kwantificeren bij ordinale en kwantitatieve variabelen</p> <p>Om bij een ordinale variabele het verschil te kwantificeren gebruik je het maximale verschil in cumulatief percentage. (max.Vcp)</p> <p>Om bij een kwantitatieve variabele het verschil te kwantificeren gebruik je de effectgrootte. Je kunt ook de boxplots vergelijken.</p> <p>De vuistregels staan op het formuleblad.</p>	<p>Het berekenen van max.Vcp</p> <ul style="list-style-type: none"> ✓ Bereken voor de variabele het cumulatieve percentage. ✓ Bereken bij elke categorie het verschil van de cumulatieve percentages. ✓ Kijk wat het grootste verschil is dat optreedt, dat is de max.Vcp <p>Zie bladzijde 66 voor een voorbeeld.</p>																

Het berekenen van de effectgrootte

Je kunt voor beide groepen het gemiddelde en de standaardafwijking berekenen. Je krijgt dan \bar{X}_1 , \bar{X}_2 , S_1 en S_2 . De effectgrootte E is dan gelijk aan:

$$E = \frac{\bar{X}_1 - \bar{X}_2}{\frac{1}{2}(S_1 + S_2)}$$

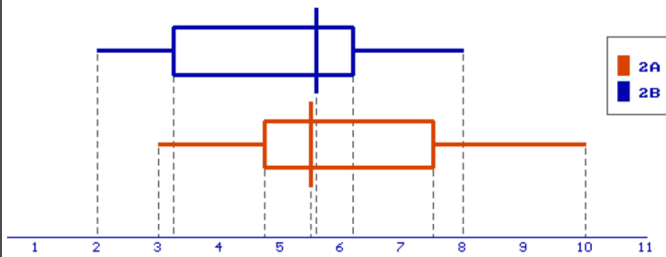
Neem $\bar{X}_1 > \bar{X}_2$ zodat de effectgrootte een positief getal oplevert.

Vuistregels bij effectgrootte

effectgrootte	het verschil is
$E > 0,8$	groot
$0,4 < E \leq 0,8$	middelmatig
$E \leq 0,4$	gering

De vuistregels staan op het formuleblad.

Boxplots vergelijken



Om boxplots te vergelijken gebruiken we de vuistregels voor het vergelijken van boxplots...:-)

Vuistregels vergelijking van boxplots

twee boxplots vergelijken	het verschil is
de boxen overlappen elkaar niet	groot
de boxen overlappen elkaar en een mediaan van een boxplot ligt buiten de box van de andere boxplot	middelmatig
alle andere gevallen	gering

De vuistregels staan op het formuleblad.

 [formuleblad](#)

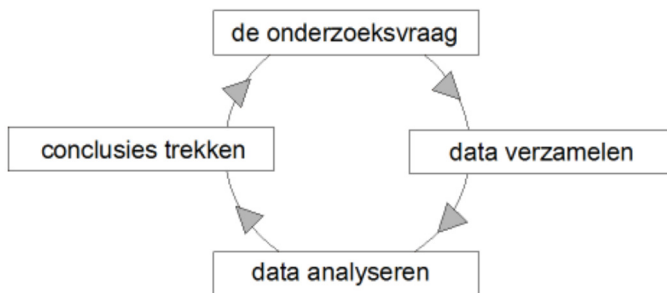
3. de statistische cyclus

Fasen in de statistische cyclus

In een deugdelijk statistisch onderzoek zijn in ieder geval de volgende fasen te onderscheiden:

1. de onderzoeksvraag
2. data verzamelen
3. data analyseren
4. conclusies trekken

Na het trekken van conclusies kan het zijn dat opnieuw een of meer fasen moeten worden doorlopen.



Daarom spreekt men van de **statistische cyclus**.

Data analyseren

Om je gegevens te analyseren kun je kijken naar de kerngegevens zoals centrum- en spreidingsmaten. Je kunt diagrammen tekenen, betrouwbaarheidsintervallen berekenen, kijken naar associatiematen en nog veel meer...



De onderzoeksvraag

Deze fase omvat de **probleemstelling** en de hierbij behorende **hoofdvraag**. Bij de hoofdvraag worden vaak **deelvragen** geformuleerd die helpen het antwoord te geven op de hoofdvraag.

In deze fase wordt ook nagedacht over de **variabelen** die een rol spelen en hoe de **gegevens** verzameld moeten worden.

Bovendien is er aandacht voor het nut van het onderzoek en de haalbaarheid.

Data verzamelen

Gegevens over Nederland kan je o.a. vinden op de website van het Centraal Bureau voor de Statistiek (CBS).

Heb je geen gegevens voor je onderzoek dan kan je (bijvoorbeeld) door een **enquête** aan gegevens proberen te komen.

Bij een enquête neem je een steekproef uit de populatie. Het is belangrijk dat de steekproef representatief is, dat wil zeggen dat de steekproef een goede weergave is van de totale populatie.

De **representatieve steekproef** is voldoende groot en aselekt.

Conclusies trekken

Kun je uit de gegevens die je hebt verzameld en de data die je hebt geanalyseerd een conclusie trekken?

Wat betekent het resultaat van de data-analyse voor de onderzoeksvraag? Ontbreken er wellicht gegevens? Heb je variabelen over het hoofd gezien? Is er wel sprake van een causaal verband?

Deze en andere vragen kan je proberen te beantwoorden door (een deel van) de cyclus opnieuw te doorlopen.

4. interpreteren van onderzoeksresultaten



In deze paragraaf beantwoord je enkele vragen naar aanleiding van teksten, tabellen en diagrammen uit de publicatie **Welvaart in Nederland 2014** van het Centraal Bureau voor de Statistiek.



In deze tweede editie van Welvaart in Nederland worden de meest actuele landelijke en regionale cijfers gepresenteerd over inkomens, vermogens en bestedingen van huishoudens. Waar mogelijk komen lange trends van deze aspecten van welvaart aan bod. Ook wordt ingegaan op de economische verwachtingen van consumenten en de druk van de inkomstenbelasting op het inkomen. Daarnaast zijn de groepen aan de onderkant en de bovenkant van de welvaarts ladder in kaart gebracht.

Speciale aandacht is er in deze editie voor het (pensioen)inkomen van ouderen, de inkomensontwikkeling van zelfstandigen met en zonder personeel en de koopkrachtontwikkeling van werknemers in de verschillende bedrijfstakken. Verder is onderzocht wie er het meest profijt hebben gehad van de verschillende toeslagregelingen.

Welvaart in Nederland 2014. Inkomen, vermogen en bestedingen van huishoudens en personen.

✓  **Welvaart in Nederland 2014**

Willems