

Wiskunde A

Getal & Ruimte

VWO 4

Samenvatting hoofdstuk 2

november 2023

Inhoudsopgave

2.1

2.2

2.3

2.4

2.5

2.1 de statische cyclus

Fasen in de statische cyclus

Bij een statisch onderzoek probeert men antwoord op een vraag te vinden door het verzamelen, verwerken en interpreteren van gegevens.

De statistiek houdt zich bezig met methoden en technieken voor het verzamelen, verwerken, analyseren en presenteren van gegevens.

- verzamelde gegevens heten **data**
- **data** staat centraal in de **statische cyclus**: het proces van opstellen van een onderzoeksvraag, verzamelen en analyseren van data en hieruit je conclusies trekken.
- **statische cyclus**: de cyclus betekent dat er als je conclusies trekt, er vaak weer nieuwe vragen ontstaan.

De volgorde voor zo'n cyclus is als volgt

1. onderzoeksvraag;
2. data verzamelen;
3. data analyseren;
4. conclusies trekken;
5. Nieuwe onderzoeksvraag? Begin bij 1.

De onderzoeksvraag

Als je gaat onderzoeken heb je een onderzoeksvraag nodig, deze is van cruciaal belang. Een goede onderzoeksvraag kan je helpen om je onderzoek gedegen uit te voeren en een overzichtelijk verslag te maken.

- Zo'n verband kan *samenhang* aangeven
VB: In welke leeftijdscategorie wordt er door jongeren het meest gesport?
- Het kan ook *causaal* zijn, dit noemen we een **causaal verband** of **oorzakelijk verband**. Een gebeurtenis is causaal als een gebeurtenis een direct gevolg is van een andere gebeurtenis.
VB: Hoe beïnvloedt slaapgedrag je concentratievermogen tijdens de les?

Stappenplan bij het opstellen van een onderzoeksvraag

1. Het onderzoeksterrein *verkennen*
Om een goede onderzoeksvraag te stellen, moet je genoeg over een onderwerp weten.
2. Het onderzoeksterrein *afbakenen*
Beperk je bij het onderzoek tot iets wat je overziet. Maak je onderzoeksterrein niet te breed.

3. De eerste versie van de onderzoeksvraag formuleren
Nadat je je onderzoeksterrein genoeg hebt afgebakend, blijven er nog verschillende soorten vragen over die je kunt stellen.
4. De onderzoeksvraag toetsen
Een onderzoeksvraag moet aan de volgende eisen voldoen:
 - a. De onderzoeksvraag moet *scherp geformuleerd* zijn
Er mogen dus *géén* vage woorden in zitten.
 - b. De onderzoeksvraag moet *enkelvoudig* zijn
Aan het woordje 'en' in een onderzoeksvraag kun je zien dat het om twee onderzoeksvragen gaat.
 - c. Een onderzoeksvraag moet *nieuw* zijn
Het heeft geen zin om onderzoek te doen naar iets wat al algemeen bekend is.
 - d. De onderzoeksvraag mag *niet uitgaan van foute veronderstellingen*
Een goede onderzoeksvraag mag er niet van uitgaan dat iets zo is, zonder daar nog onderzoek naar te hebben gedaan.

Een steekproef nemen

Soms kan er niet bij iedereen een onderzoek gedaan worden, omdat de omvang van het onderzoek anders te groot wordt. Daarom worden er **steekproeven** uitgevoerd.

Voorwaarden bij een steekproef

- Een statistisch onderzoek hoeft zich niet alleen te richten op individuen, maar mag ook betrekking hebben op huishoudens, pakken melk, scholen, lampen, etc.
- De totale groep waarop het onderzoek zich richt heet de **populatie**.
- De steekproef moet **representatief zijn**, verhoudingen tussen groepen worden in verhouding gehouden.
- De steekproef moet **voldoende groot** zijn om representatief te zijn. Hoe meer kenmerken je onderzoekt, hoe groter de steekproef.
- De steekproef is **aselect**, dit betekent dat elk element van de populatie evenveel kans heeft om in de steekproef voor te komen.

2.2 Data verzamelen

Enquêtevragen en statistische variabelen

In dit paragraaf wordt ingezoomd op het 2^e deel van de statische cyclus: data verzamelen. Om onderzoek te krijgen op je onderzoeksvraag moeten er gegevens verzameld worden.

Binnen de statistiek wordt er onderscheid gemaakt tussen twee soorten variabelen.

Kwalitatieve variabelen	geven een kwaliteit of eigenschap weer VB: beroep of godsdienst
<i>discrete variabele</i>	is een variabele die alleen bestaat uit helen
<i>continue variabele</i>	is een variabele die uit alle getallen kan bestaan

Kwantitatieve variabelen	geven meetbare gegevens weer VB: temperatuur en snelheid
---------------------------------	---

Meetniveaus

De nominale variabele is op nominaal niveau gemeten, hierbij maakt het niet uit in welke volgorde je de variabelen zet.

De ordinale variabele is op ordinaal niveau gemeten, hierbij maakt het wél uit in welke volgorde je de variabelen zet.

Behalve de nominale en de ordinale variabele zijn er nog twee andere meetniveaus. Nummer 3 & 4.

1. Nominaal meetniveau
 - er wordt alleen onderscheid gemaakt tussen verschillende categorieën, meestal aangegeven met de namen. Volgorde maakt niet uit
2. Ordinaal meetniveau
 - er wordt onderscheid gemaakt tussen verschillende categorieën
 - er wordt een ordening (volgorde) aan gegeven
3. Interval meetniveau
 - er wordt onderscheid gemaakt tussen verschillende categorieën
 - er wordt een ordening (volgorde) aan gegeven
 - de verschillen tussen de variabelen zijn gelijk
4. Ratio meetniveau
 - er wordt onderscheid gemaakt tussen verschillende categorieën
 - er wordt een ordening (volgorde) aan gegeven
 - de verschillen tussen de variabelen zijn gelijk
 - er is een natuurlijk nulpunt

Populatiegrootte en steekproefproportie

Er wordt een onderzoek gedaan. Bekend is dat 7,9% van de onderzoekskandidaten in aanmerking komen voor optie A. De populatiegrootte is dan 0,079, dit geven we aan met de letter p . De populatiegrootte is dus de grootte van een populatie die een bepaald kenmerk heeft, uitgedrukt in een decimaal.

Bij een steekproef van 2000 mensen blijken 250 in aanmerking te komen voor optie A. De steekproefproportie is dan 0,125. Want $steekproefproportie = \frac{2000}{250} = 0,125$

2.3 Centrum- en spreidingsmaten

In deze paragraaf gaat het over fase 3 van de statische cyclus: data analyseren.

Centrummaten

Bij het analyseren van data kan het nuttig zijn om een serie waarnemingsgetallen te typeren voor één getal. Daar neem je vaak het gemiddelde voor.

Maar het gemiddelde geeft niet altijd de gewenste indruk, uitschieters kunnen vaak het gemiddelde flink beïnvloeden. Je kunt in dat geval beter gebruik maken van de mediaan of modus.

De gemiddelde, mediaan en modus zijn centrummaten.

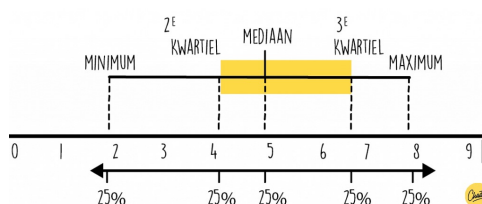
centrummaat	uitleg	voordeel	nadeel
gemiddelde	Het gemiddelde is $\frac{\text{som van de waarnemingsgetallen}}{\text{totale frequentie}}$	Alle gegevens worden gebruikt	Gevoelig voor uitschieters
mediaan	- Schrijf de getallen in volgorde van grootte - De mediaan is het middelste getal bij een oneven aantal getallen Bij een even aantal zijn het de twee middelste	Niet gevoelig voor uitschieters	Alleen de grootte van de middelste waarnemingen is van belang
modus	Waarnemingsgetal met de grootste frequentie, als er twee of meer getallen zijn met dezelfde hoogste frequentie, is er geen modus.	De enige centrummaat die bij kwalitatieve gegevens te gebruiken is	Is niet altijd aanwezig Een kleine verandering kan een geheel andere modus opleveren

Boxplot

Een boxplot is een diagram dat de verdeling van data in beeld brengt.

Alleen bij een kwantitatieve variabele kan een boxplot gemaakt worden.

Een boxplot neemt de mediaan als “middelpunt”, of ja, middelste verdelingslijn. Van de twee vlakken die er gemaakt worden kan er weer de mediaan worden genomen en dan ontstaat er links van de mediaan (het eerste kwartiel) de Q1 en rechts (het derde kwartiel) de Q3.



voorbeeld van een boxplot mr-chadd.nl

De standaardafwijking

Twee groepen met een gelijke mediaan kunnen nog steeds sterk van elkaar verschillen. Om nog meer informatie te geven over een groep, kun je gebruik maken van de **spreidingsmaat**. De spreidingsmaat drukt de mate van spreiding uit in een getal.

Spreidingsbreedte de eenvoudigste spreidingsmaat, maar zeer gevoelig voor uitschieters. Hoe breed een boxplot spreidt.

Interkwartielafstand de afstand per kwartiel

Het symbool van de standaardnotatie σ , de Griekse letter sigma.

De standaardafwijking met de hand berekening is heel moeilijk, daarom doe je het via je GR. Je berekent 'm zo:

1. bereken het gemiddelde van de waarnemingsgetallen
2. bereken voor elk waarnemingsgetal het verschil met het gemiddelde, dus *waarnemingsgetal – gemiddelde*.
3. tel de kwadraten van al deze verschillen bij elkaar op.
4. deel de uitkomst door het aantal waarnemingsgetallen

2.4 Klassenindeling en polygonen

Dit paragraaf gaat ook over fase 3, Data analyseren.

Klassenindeling

Als je heel veel verschillende waarden hebt, zoals

55	28	36	68	21	17	65
32	20	26	15	58	11	43
57	69	33	27	47	38	56
29	25	14	18	16	51	44

heeft het geen zin om een frequentietabel te maken, want elke waarde komt maar één keer voor. Het is nu beter om de waarden in klassen in te delen, deze groepen heten **klassen**.

Omdat het kleinste bedrag 11 euro is en het grootste 69 euro, is een verstandige klassenindeling $[10, 20>$ $[20, 30>$ $[30, 40>$ $[40, 50>$ $[50, 60>$ $[60, 70>$

$[$ = het getal hoort er bij

$<$ = de klasse begint ná dit getal

$]$ = tot en met dit getal

$>$ = tot dit getal

De klassenbreedte bereken je door het laatste getal van de klasse te verminderen met de eerste, dus in het geval van de klasse $[10, 20>$ is dat $20 - 10 = 10$.

Het klassenmidden bereken je door de waarden van je klasse bij elkaar op te tellen en dit te delen door 2, dus in het geval van de klasse $[20, 30>$ is dat $\frac{20+30}{2} = 25$

Als je een frequentietabel staat kun je niet meer zien van welke waarnemingsgetallen is uitgegaan, bij de klasse $[10, 20>$ kan zeven keer de drie en acht keer de vijf zitten maar ook vijftien keer de negen. Om de nadering te krijgen van het gemiddelde gebruik je daarom de klassenmiddens.

In plaats van de modus spreken we van de modale klasse bij een klassenindeling.

Cumulatieve frequentiepolygoon

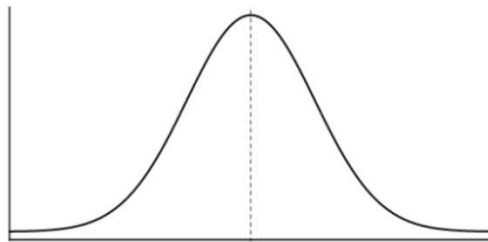
Tel je bij de frequentie van een waarnemingsgetal ook alle voorgaande frequenties mee, dan spreken we van de **cumulatieve frequentie**. Wanneer je met cumulatieve frequenties in procenten te maken hebt, noem je het resultaat de **relatieve cumulatieve frequentie**.

2.5 Soorten verdelingen

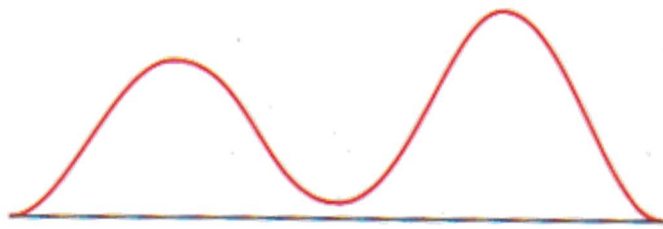
Verdelingen en histogrammen

Bij een symmetrische verdeling vallen het gemiddelde, de mediaan en de modus samen.

Symmetrische verdeling

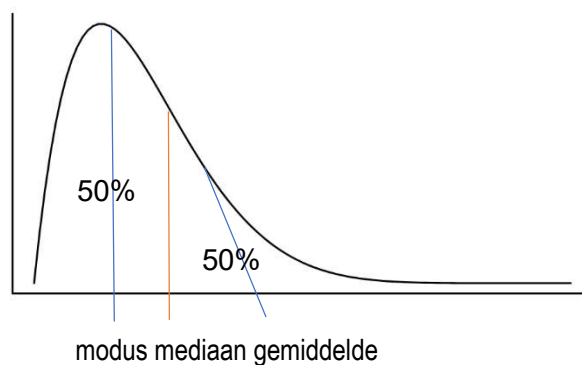


Tweetoppige verdeling



Rechts-scheve verdeling, hierbij geldt modus, mediaan en dan gemiddelde

Rechts-scheve verdeling



Links van de mediaan ligt 50% en rechts van de mediaan 50%. Voor de links-scheve verdeling geldt het gespiegelde van deze afbeelding.

Links-scheve verdeling, een gespiegelde rechts-scheve verdeling

Links-scheve verdeling

