

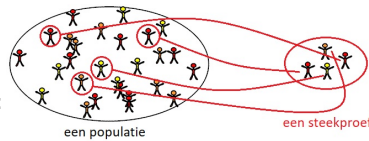
Spiekbriefjes bij Beschrijvende statistiek

Metingen en statistiek

Het verzamelen en presenteren van gegevens heet **beschrijvende statistiek**.

Een **statistisch onderzoek** kent diverse fasen:

- Je begint met een **onderzoeksvraag**.
- Je bepaalt de **populatie** (doelgroep) en neemt daaruit een **steekproef**. Deze steekproef moet **representatief** (kenmerkend) zijn en niet te klein. Hij is **aselect**, elk lid van de populatie heeft dezelfde kans om in de steekproef te komen.
- Je stelt vast over welke **statistische variabele** het gaat. Er zijn **kwantitatieve variabelen** met getallen als waarden en **kwalitatieve variabelen** met eigenschappen als waarden.
- Je verzamelt **data** (statistische gegevens) die je gaat **ordenen** middels tabellen en diagrammen: tabellen zijn nauwkeuriger, diagrammen overzichtelijker.
- Je probeert een (eerlijke) **conclusie** te trekken.



meer info

Meetnauwkeurigheid

Bij metingen wordt worden **meetfouten** of **meetnauwkeurigheden** gemaakt:

- Een **persoonlijke fout** is een fout die je zelf maakt. Bijvoorbeeld een afleesfout. Schat van meetwaarden alleen het laatste cijfer!
- Een onvermijdelijke **toevallige fout**. Bij een meetinstrument van meetklasse 2,0 is er een maximale afwijking van 2% van de schaallengte (of meetbreedte). Dit betreft de **precisie** van de metingen.
- Een **systematische fout** geeft een afwijking in één richting, bijvoorbeeld omdat je meetinstrument niet geijkt is. Of je leest systematisch scheef af, er is sprake van **parallax**. Het betreft **juistheid** van de metingen.

Meet je $20 \pm 0,5$, dan is 0,5 de **absolute fout**.

Deel je de meetnauwkeurigheid door de hoeveelheid, dan krijg je de **re-**

latieve fout. Hier $\frac{0,5}{20} = 0,025$, ofwel 2,5%.



meer info

Diagrammen

Bij veel statistische variabelen past een **frequentietabel**. Hoe vaak een waarde voorkomt is de **absolute frequentie**.

Deel je die door het totale aantal, dan krijg je de **relatieve frequentie**. Relatieve frequenties (vaak procenten) maken vergelijken van datasets makkelijker.

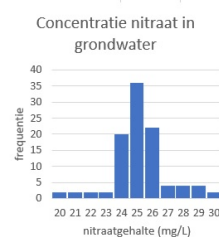
Vaak gebruik je een **klassenindeling**. Bij metingen horen vanwege de meetnauwkeurigheid altijd klassen.

In een tabel heb je de data nauwkeurig voorhanden, een **diagram** geeft meer overzicht:

Bij één variabele gebruik je een **beelddiagram**, een **staafdiagram**, een **lijndiagram**, een **cirkeldiagram**, een **steelbladdiagram**, een **kaartdiagram**, of...

Om twee variabelen te vergelijken gebruik je een **puntenwolk**.

Concentratie nitraat (mg/L)		
klasse	freq	rel.freq
20	1	2
21	1	2
22	1	2
23	1	2
24	10	20
25	18	36
26	11	22
27	2	4
28	2	4
29	2	4
30	1	2
	50	100

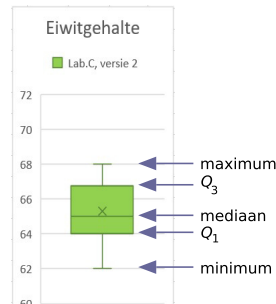


meer info

Centrummaten

Kwantitatieve gegevens vat je samen met **centrummaten**:

- de **modus** (modale klasse): meest voorkomende waarde;
- de **mediaan**: de middelste waarde met de gegevens op volgorde van grootte;
- het **gemiddelde**: de waarden opgeteld en gedeeld door het totaal aantal.



Bij populatiegemiddelde μ is de **juistheid** $|\bar{x} - \mu|$ en de relatieve juistheid $\frac{|\bar{x} - \mu|}{\mu} \cdot 100\%$.

In een **boxplot** is de mediaan het midden, het eerste kwartiel Q_1 de mediaan van de eerste helft en het derde kwartiel Q_3 de mediaan van de tweede helft. Je kunt zo datasets vergelijken.

Bij kwalitatieve data hebben centrummaten vaak geen betekenis.



meer info

Spreidingsmaten

Kwantitatieve statistische gegevens hebben ook **spreidingsmaten**:

- de **spreidingsbreedte**: het verschil tussen de hoogste en de laagste waarde;
- de **(inter)kwartielafstand** in een boxplot: $IQR = Q_3 - Q_1$;
- de **gemiddelde absolute afwijking** (GAA): het gemiddelde van alle positieve afwijkingen van het gemiddelde, $GAA = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$.

De **precisie** is de relatieve spreidingsbreedte, de spreidingsbreedte gedeeld door het gemiddelde en omgerekend naar procenten. Er zijn ook andere maten voor de precisie.

Bij kwalitatieve data hebben spreidingsmaten vaak geen betekenis.

Soms is er een **uitschieter** (of uitbijter), een waarde die erg veel afwijkt.

- In een boxplot zit een uitschieter meer dan $1,5 \times IQR$ onder het eerste of boven het derde kwartiel.
- Bij kleine steekproeven kun je (soms) **Dixon's Q-test** toepassen.



meer info