

Om een goed onderzoek te doen gebruik je 4 fases:

1. Onderzoeksfase
2. Data verzamelen
3. Data analyseren
4. Conclusie trekken

Bij een onderzoek zoek je ook naar een verband tussen 2 dingen, vind je dit en heeft dit een directe reden tussen mekaar. Dan spreek je over een oorzakelijk of causaal verband.

Om een onderzoeksvraag op te stellen ga je door 4 fases:

- Verkennen
- Afbakenen
- Eerste versie formuleren
- Toetsen

Let bij het toetsen op de volgende dingen:

- Scherp geformuleerd
- Enkelvoudig
- Nieuw
- Niet uit gaan van een foutieve veronderstellingen

Als je achter gegevens wilt komen moeten deze ergens vandaan komen. Dit doe je vaak doormiddel van een steekproef. Die moet voldoende groot zijn en aselctief zijn. Dit zorgt voor een goede afspiegeling van het geheel.

2.2

Als je deelnemers voor een enquête hebt gevonden heten dit respondenten.

Je data bestaat uit allerlei variabelen. Dit zijn dan weer kleinere vraagjes of onderzoekjes.

De variabelen kun je weer opdelen in:

- Kwalitatieve variabelen (bijv, een rugnummer bij voetbal)
- Kwantitatieve variabelen (bijv, temperatuur)
 - Discrete variabele (bijv, schoenmaat)
 - Continue variabele (bijv, gewicht)

Je hebt verschillende meetniveaus:

- Nominaal, Slechts onderscheid makend in categorien. Zoals: naam, beroep, woonplaats en geslacht.
- Ordinaal, Om een volgorde te bepalen. Zoals: het aantal sterren of het rangensysteem in de krijgsmacht.
- Interval, Volgorde + gelijke verschillen zonder natuurlijk nulpunt. Zoals: de jaartelling.
- Ratio, Volgorde + gelijke verschillen met natuurlijk nulpunt. Zoals: je leeftijd.

Om duidelijkheid te kunnen geven bij een steekproef bestaat de steekproefproportie. Die bereken je door:

$$p = \frac{\text{Aantal elementen}}{\text{Totaal}}$$

2.3

Je hebt 3 centrummaten. Nml:

- Gemiddelde
- Mediaan
- Modus

Om centrummaten overzichtelijker te maken heb je de boxplot. Die is opgedeeld in delen van 25%. Die delen heten kwartielen.

Ken deze formule:

$$\text{spreidingsbreedte} = \text{maximum} - \text{minimum}$$

en

$$\text{interkwartielafstand} = Q3 - Q1$$

Je moet de standaardafwijking (σ) kunnen berekenen. Dit kan door:

- Gemiddelde berekenen
- Voor elk geval verschil met gemiddelde berekenen
- Tel de kwadraten bij elkaar op
- Deel uitkomst door aantal waarnemingsgetallen
- Doe de wortel van de uitkomst

2.4

Om groepen duidelijker te maken, maak je gebruik van

- Klassengrenzen, deze kan er als volgt uit zien: $[10,20>$
 - Dit betekent dat 10 WEL mee doet en 20 NIET mee doet.
- De klassenbreedte is hoeveel er in de klassen valt, dus: $20-10=10$
- Het klassenmidden is het gemiddelde van de klasse, dus: $(10+20):2=15$

De waarnemingsgetallen worden onduidelijk in een klassenindeling. Daarom kun je klassenmiddens gebruiken (frequentie).

In plaats van te spreken van de modus spreek je bij klassenindelingen van de modale klasse.

In een histogram (lijkt op staafdiagram) kun je de frequentieverdeling invullen. Maar ook andersom.

Als je een lijndiagram maakt van de frequentieverdeling heet dit een frequentiepolygoon. De gegevens zijn af te lezen aan de puntjes.

Er zit een verschil in de relatieve en absolute frequentiepolygoon. Bij relatieve gebruik je percentages, terwijl je bij de absolute frequentiepolygoon de "normale" getallen gebruikt.

Dan heb je nog de cumulatieve frequentiepolygoon, hierbij tel je de frequenties bij elkaar op. Je krijgt ene stijgende lijn

2.5

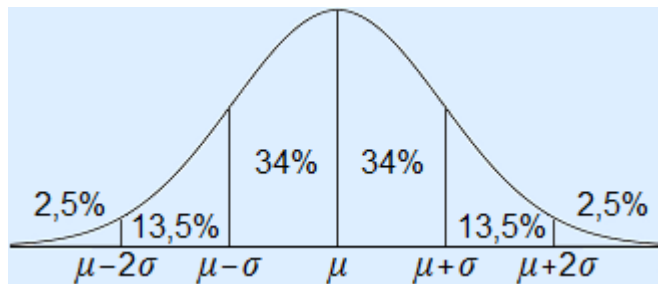
Als je alle data in een histogram hebt gezet kun je het gaan plaatsen in een van de volgende verdelingen:

- Symmetrische verdeling (2 kanten gelijk)
- Tweetoppige verdeling (heeft 2 toppen)
- Rechts-scheve verdeling (aan de rechter kant een staart)
- Links-scheve verdeling (aan de linkerkant een staart)
- Uniforme verdeling (is recht)

Je moet deze verdelingen ook kunnen toepassen in een boxplot. Visualiseer het hiervoor voor in een verdeling. Dit kan het makkelijker maken.

Je moet ook een cumulatieve verdelingskromme kunnen maken. Dit lijkt een beetje op de vloeiende lijn van de histogram. Alleen dan vloeiend. Ook de verdelingskromme is zowel cumulatief als relatief te vinden. Wederom met hetzelfde verschil. De cumulatieve kromme betreft een kromme met daarin een schets of een reeks waarnemingsgetallen en bij een relatieve kromme word alles aangegeven met percentages.

Dan de normale verdeling. Die ziet er als volgt uit:



Hier is te zien dat 68% ligt tussen $\mu - \sigma$ en $\mu + \sigma$

Hier is ook te zien dat 95% ligt tussen $\mu - 2\sigma$ en $\mu + 2\sigma$

2.6

Je moet de standaardafwijking (σ) kunnen berekenen. Dit doe je als volgt:

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

P = Het gemiddelde (deel:totaal=p)

n = Het totaal

De volledige formule voor het 95% betrouwbaarheidsinterval is:

$$[p - 2\sigma, p + 2\sigma]$$

Maar die krijg je gegeven op het proefwerk.

Voor het populatiegemiddelde ziet die er echter ietsjes ander uit namelijk:

$$\bar{x} - 2 \cdot \frac{s}{\sqrt{N}}, \bar{x} + 2 \cdot \frac{s}{\sqrt{N}}$$