

Inhoudsopgave: Datum:

# Hoofdstuk 1: Basic Concepts and Sampling

## Basic Concepts and Sampling

**Variabelen en waarden:**

* **Variabele**
  + Algemene eigenschap van een object, maakt het mogelijk om objecten te onderscheiden
* **Waarde**
  + Specifieke eigenschap, interpretatie voor die variabele

Measurement Levels:

* Typen variabelen
* Meest geschikte methode voor analyse bepalen
  + visualisatiemethoden
  + centrale tendens en spreiding
  + relatie tussen variabelen onderzoeken

Qualitative vs quantitative:

* **Kwalitatief** 
  + Niet noodzakelijkerwijs numeriek
  + Beperkt aantal waarden
* **Kwantitatief**
  + Getal + meeteenheid
  + Veel waarden, vaak uniek

Qualitative scales:

* **Nominaal** 
  + Dit zijn de categorieën. Dit zijn ook kwalitatieve eigenschappen. Geen rangschikking
    - geslacht
    - ras,
    - land
    - vorm, ...
* **Ordinaal** 
  + Rangorde. Dit zijn de kwantitatieve eigenschappen. Wel een rangschikking
    - militaire rang,
    - opleidingsniveau, ...

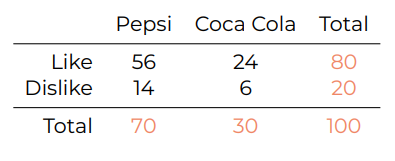
Quantitative scales:

* **Interval** 
  + Geen vast nulpunt ⇒ geen verhoudingen
  + Voorbeeld: °C, °F
* Ratio
  + Een absoluut nulpunt => proporties
  + Voorbeeld: afstand (m), energy (j), gewicht (kg)
* **Verhouding** 
  + Absoluut nulpunt ⇒ verhoudingen
  + Voorbeeld afstand (m), energie (J), gewicht (kg) ...
* **Verhoudingen**:
  + 20 m is 1/3e of ongeveer 33% langer dan 15 m
  + 20 °C is NIET 1/3e warmer dan 15 °C (omrekenen naar °F)

Afbeelding met lijn, diagram, Perceel, tekst

Door AI gegenereerde inhoud is mogelijk onjuist.Relations between variables:

Variabelen zijn gerelateerd als hun waarden **systematisch** veranderen.

Relations between variables example:

* Is er een **verband** tussen het type cola en de smaakwaardering?
  + Ja: de **verhoudingen hier zijn hetzelfde**: 14/56 = 25% en 6/24 is ook 25%

Causal Relationships:

* Onderzoekers zijn vaak op zoek naar **causale verbanden**, bijv.
  + Frustratie leidt tot agressie
  + Alcohol leidt tot verminderde alertheid
* **Oorzaak** 
  + Onafhankelijke variabele
* **Gevolg** 
  + Afhankelijke variabele

Fake correlations or “Spurious correlations”:

* **Waarschuwing**
  + Een verband tussen variabelen duidt niet noodzakelijk op een oorzakelijk verband!
* **Voorbeelden**:
  + Gewelddadige videospelletjes leiden tot gewelddadig gedrag
  + Vaccins kunnen autisme veroorzaken
  + Relatie tussen cola light drinken en obesitas

## Sample testing

Afbeelding met tekst, cirkel, Lettertype, logo

Door AI gegenereerde inhoud is mogelijk onjuist.Suppose you want to analyze a group of friends:

* Vragen die je kunt stellen:
  + Hoe lang zijn mijn vrienden?
  + Wat zijn hun gewichten?
  + Hoe veilig is hun leefomgeving?
  + Hebben ze familie? ...

Sample and Population:

* **Populatie** de verzameling van alle objecten/mensen/... die je wilt onderzoeken
* Afbeelding met Fictief personage, Superheld, Held, tekenfilm

  Door AI gegenereerde inhoud is mogelijk onjuist.**Steekproef een deelverzameling** van de populatie waaruit metingen worden gedaan
* Onder bepaalde omstandigheden zijn de resultaten van een steekproef representatief voor de populatie
* Een steekproef is gemakkelijker te analyseren dan de hele populatie

Afbeelding met tekst, schermopname, Lettertype, lijn

Door AI gegenereerde inhoud is mogelijk onjuist.Sampling method (aan te vullen in de les):

How to pick elements for a sample?

* **Random steekproef**: elk element uit de populatie heeft een gelijke kans om in de steekproef opgenomen te worden.
* **Niet-random steekproef**: de elementen voor de steekproef zijn niet willekeurig gekozen. Objecten die gemakkelijk kunnen worden verzameld, hebben meer kans om te worden opgenomen (gemakssteekproef).

Afbeelding met tekst, schermopname, Lettertype, nummer

Door AI gegenereerde inhoud is mogelijk onjuist.Stratified to variables (aan te vullen in de les):

Possible Errors:

* Metingen in een steekproef zullen meestal afwijken van de waarde in de hele populatie ⇒ **Errors**!
* Toevallig ↔ Systematisch
* Steekproeffout ↔ Niet-steekproeffout

Sampling Errors :

* **Toevallige steekproeffouten** 
  + Puur toeval – bv. je neemt een aantal studenten, maar neemt per ongeluk allemaal jongens
* **Systematische steekproeffouten** 
  + **Online enquête**: mensen zonder internet zijn uitgesloten
  + **Straatonderzoek**: alleen wie er op dat moment loopt
  + **Vrijwillige enquête**: alleen geïnteresseerden nemen deel

Non-sampling Errors:

* **Perongelijke niet-steekproeffouten** 
  + Verkeerd aangekruiste antwoorden
* **Systematische niet-steekproeffouten** 
  + **Slechte** of niet-gekalibreerde **meetapparatuur**
  + **Waarde** kan beïnvloed worden door het feit dat je meet
  + **Respondenten liegen** (aantal sigaretten per dag)

# Hoofdstuk 2: Univariate statistics

## Central Tendency and Dispersion

How tall are my friends?

Welke waarde is representatief voor de hele groep?

Mean or avarage:

* Afbeelding met Lettertype, lijn, nummer, ontwerp

  Door AI gegenereerde inhoud is mogelijk onjuist.**Rekenkundig gemiddelde**
  + Het rekenkundig gemiddelde (notatie: 𝑥) is de som van alle waarden gedeeld door het aantal waarden
* Q1 Wat gebeurt er als Ant-Man krimpt tot een grootte van 10 cm?
* Q2 Het rekenkundig gemiddelde van 15 getallen is 12. Welk getal moet erbij opgeteld worden om een gemiddelde van 13 te krijgen?

Mediaan:

* Om de mediaan te vinden, sorteer je alle waarden en kies je het middelste getal.
  + Oneven aantal waarden: geen probleem
  + Even aantal waarden: gemiddelde van de middelste twee
* Q1 Wat gebeurt er als Ant-Man krimpt tot een grootte van 10 cm?
* Q2 Wat is de mediaan van het aantal mensen dat gered is door Batman in de afgelopen acht jaar?

## Data Visualisation