# HT 1

Gegevens in een datamatrix verzamelen

Rijen (waarnemingen) Kolommen (variabelen) Cellen (waarden)

Exploratie van de datamatrix: aandacht voor meetniveau, extreme waarden, ontbrekende records en waarden

Extreme waarden: fouten?

Soort analyse is afhankelijk van het meetniveau van de variabele!

# HT 2

Centrum en spreiding: belangrijk om inzicht te krijgen in **variabele**

* Maten voor centrum en spreiding: waar is zwaartepunt, hoe zijn waarnemingen verdeeld rond dat zwaartepunt
* Vorm van een verdeling aan de hand van histogram

🡪 Verschaffen inzicht, maar zijn op zich geen ‘antwoord’

|  |  |  |
| --- | --- | --- |
| MEETNIVEAU | CENTRUMMATEN | SPREIDINGSMATEN |
| Nominaal | Modus | / |
| Ordinaal | Modus, Mediaan | Variatiebreedte, Interkwartielafstand |
| Kardinaal | Modus, Mediaan, Gemiddelde | Variatiebreedte, Interkwartielafstand,  Standaardafwijking,Variatiecoëfficiënt |

# HT 3

1. Statistisch verband? Nulhypothese en alternatieve hypothese
2. Voorstellen verband tussen nominale variabelen: kruistabellen, stapeldiagram, geclusterd staafdiagram
3. Statistisch verband tussen nominale variabelen: associatiematen

* Chi-kwadraat: meet verschil tussen verwachte en waargenomen waarden
* Voorwaarden voor een geldige chi-kwadraat test
  + - Statistische significantie chi-kwadraat uitkomst: p-waarde vergelijken met vooropgestelde α-waarde
    - H0 aanvaarden: p > 0,05 (geen stat. significant verband)
    - H0 verwerpen: p < 0,05 (kans klein dat het stat. Verband toevallig is)
* Sterkte van het verband: Cramers V

🡪 Significant ≠ Relevant

# HT 4



# HT 5

|  |  |  |
| --- | --- | --- |
| **Meetniveau** | **Soort analyse** | **Statische maten** |
| Nominaal | Kruistabellen  Associatie | c2 (chi-kwadraat)  Cramer’s V |
| Ordinaal | Rangorde-correlatie | Spearmans Rho  Kendall tau |
| Kardinaal | Correlatie (niet-causaal) | Pearson’s R  (Correlatiecoëfficiënten) |
| Regressie (causaal verband) | Regressiecoëfficiënten in de vergelijking  (Y = a + b\*x)  a (= intercept, snijpunt met Y-as)  b (= richtingcoëfficiënt)  R2determinatiecoëfficiënt |

# HT 6

Voorstelling tijdreeksen:

Grafisch: X-as = tijd (tijdsinterval respecteren)

Y-as = variabele (schaal: metrisch of logaritmisch)

|  |  |
| --- | --- |
| **Maatstaven** | **Grafische weergave** |
| 1. Groeicijfers  -Gemiddelde groei  -Groeivoet  -Gemiddelde groeivoet | (Lijndiagram) |
| 2. Lineaire trend | Rechte trendlijn |
| 3. Gemiddelden  -Voortschrijdend  -Periodiek | Golvende trendlijn  Getrapte curve |
| 4. Indices  -Enkelvoudig  -Samengesteld (gewogen/ongewogen) | Stralendiagram |

# HT 7

*Steekproefgemiddelde*: uitkomst van één steekproef

🡪 In welke mate zegt ons dat iets over populatie?

*Betrouwbaarheidsinterval* rond steekproefgemiddelde:

* + Interval waarbinnen met …% zekerheid populatiegemiddelde ligt
  + = steekproefgemiddelde ± de *schattingsfout* (m)

Kan ook voor correlatiecoëfficiënten, regressiecoëfficiënten enz.

*Steekproef*: selectie uit grotere populatie

**Steekproefontwerp/methode:** samenstelling

-aselect en willekeurig

-representatief

Steekproefgrootte

-betrouwbaarheid

-nauwkeurig