

Formuleblad — Beschrijvende statistiek

3de jaar (5u) · centrummaten · spreidingsmaten · boxplot · correlatie

Begrippen

Populatie = volledige groep die je onderzoekt.

Steekproef = het deel dat je echt waarneemt.

Variabele = het kenmerk (krijgt waarden).

Kwantitatief (getallen): discreet (telling) / continu (meting).

Kwalitatief (eigenschap): nominaal (geen rangorde) / ordinaal (wel rangorde).

Frequenties

$af_i (n_i)$ = absolute frequentie (aantal keer).

$rf_i = af_i / n \rightarrow \text{som} = 1$; $rf\% = rf_i \cdot 100$ (som 100%).

caf_i = som van af t.e.m. die waarde ($\leq$).

$crf_i = caf_i / n$.

Centrummaten

Gemiddelde $\bar{x} = (\sum x_i \cdot n_i) / n$

(deel door n = som van de frequenties!)

Modus Mo = waarde met grootste af.

Mediaan Me = middelste waarde (50% eronder).

Mediaan & kwartielen -- stappen

1. Sorteert van klein naar groot.

2. **rangnummer Me** = $(n+1)/2$.

3. Natuurlijk getal $\rightarrow$ Me = waarde op die plaats.

4. Kommagetal $\rightarrow$ Me = gemiddelde van de twee burens.

Q1 = mediaan eerste helft (25% eronder).

Q2 = Me (50%). **Q3** = mediaan tweede helft (75%).

Oneven n : reken Q1 en Q3 **inclusief** de mediaan.

Spreidingsmaten

Variatiebreedte R = max – min

Interkwartielafstand IKA = $Q3 - Q1$ (middelste 50%)

Variantie $\sigma^2 = [\sum n_i (x_i - \bar{x})^2] / n$

Standaardafwijking $\sigma = \sqrt{\sigma^2}$

Grote σ = veel spreiding ; kleine σ = dicht bij gemiddelde.

Boxplot (7 kengetallen)

Min · Q1 · Me · Q3 · Max (+ R en IKA).

Rechthoek van Q1 tot Q3, streepje op Me.

Snorharen tot min en max. Sterretje op het gemiddelde.

Elk van de 4 stukken = 25%.

$> Q1$: 75% | tussen Q1 en Q3: 50% | $> Q3$: 25%.

Verdelingen

Symmetrisch: Me $\approx$ gemiddelde (midden).

Scheef: gemiddelde ligt aan de kant van de staart.

Uitschieter: ligt ver van de rest (fout of uitzondering).

Cluster: groep met gemeenschappelijk kenmerk.

Bivariaat: correlatie

Spreidingsdiagram: punten (x, y) = puntenwolk.

Regressierechte $y = ax + b$: best passende rechte (voorspellingen).

Positieve $a \rightarrow$ stijgend (positief verband); negatieve $a \rightarrow$ dalend.

Correlatiecoëfficiënt $-1 \leq r \leq 1$ (bereken met Excel).

$|r|$ dicht bij 1 = sterk ; $r \approx 0$ = geen lineair verband.

Examen-instinkers

· **Gemiddelde uit tabel**: deel door n (som frequenties), NIET door het aantal verschillende waarden.

· **Modus** = de waarde met de grootste frequentie, niet de frequentie zelf.

· **Mediaan**: eerst sorteren! Vergeet het gemiddelde van de twee burens niet bij even n .

· **Gemiddelde** is gevoelig voor uitschieters, de **mediaan** en **IKA** veel minder.

· **r blijft gelijk** als je x en y omwisselt; **r is gevoelig voor uitschieters**.

· **Misleidend**: verticale as niet op 0, ontbrekende of niet-chronologische tijdpunten.