

# Statistiek: groepen vergelijken op basis van statistische variabelen

Havo 5, wiskunde A, schooljaar 17/18

## Algemeen

Bepaal of je variabelen nominaal of ordinaal zijn. Op basis daarvan beslis je welke van de onderstaande associatiematen je gaat gebruiken.

Meetniveaus & hun kenmerken		Scale	Rationiveau
		<b>Intervalniveau</b>	Verhouding blijven gelijk
	<b>Ordinaal niveau</b>	Gelijke verschillen	Gelijke verschillen
<b>Nominale niveau</b>	Ordering	Ordering	Ordering
Onderscheid	Onderscheid	Onderscheid	Onderscheid
<i>Geslacht</i>	<i>Opleidingsniveau</i>	<i>Intelligentie</i>	<i>Leeftijd</i>

## 1. Vergelijken van groepen op basis van nominale variabelen (kwalitatief onderzoek)

De volgende associatiematen kun je alleen berekenen als je nominale variabelen in een 2x2-kruistabel verwerkt kunnen worden:

		Variabele X		
		wel	niet	
Variabele Y	wel	a	b	a+b
	niet	c	d	c+d
		a+c	b+d	totaal

Hierin zijn a, b, c en d de frequenties van de betreffende variabelen.

De te berekenen associatiematen zijn:

- PercentageVerschil** → 
$$PV = \frac{a}{a+b} \cdot 100\% - \frac{c}{c+d} \cdot 100\%$$

(Let op: altijd grootste – kleinste, dus uitkomst altijd positief)

- Odds Ratio** → 
$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

(Let op: altijd grootste waarde : kleinste waarde, dus uitkomst altijd >1)

- Phi-coëfficiënt** → 
$$phi = \frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

Er geldt nu het volgende:

PV	OR	phi	Conclusie: het verschil is
$0 < PV \leq 20\%$	$1 \leq OR < 2$	$-0,2 < phi < 0,2$	gering (of klein)
$20 < PV \leq 40\%$	$2 \leq OR \leq 3$	$-0,4 < phi < -0,2$ of $0,2 < phi < 0,4$	middelmatig
$PV > 40\%$	$OR > 3$	$phi < -0,4$ of $phi > 0,4$	groot

## 2. Vergelijken van groepen op basis van ordinale (kwantitatieve) variabelen

Indien je van twee groepen de beschikking hebt over ordinale of (opeenvolgende) kwantitatieve gegevens in een frequentietabel, dan kun je het maximale PercentageVerschil uitrekenen.

- **max  $V_{cp}$**  bereken je als volgt:
  1. Bereken per groep de cumulatieve frequentie van elke klasse
  2. Bereken per groep de relatieve cumulatieve frequentie van elke klasse
  3. Bereken het cumulatieve percentageverschil (dus verschil in relatieve cumulatieve frequenties) van elke klasse
  4. Lees het grootste verschil af; dit grootste verschil noemen we  $max V_{cp}$
  5. Voor de conclusie kun je dezelfde waarden als PV (zie vorige bladzijde) gebruiken.

Een andere mogelijkheid is om van beide groepen het gemiddelde  $\bar{X}$  en de steekproefstandaardafwijking  $S$  te bepalen (soms zijn deze al gegeven, zo niet m.b.v. lijsten op de GR en 1-varStats). Je kunt nu de effectgrootte  $E$  berekenen:

- **Effectgrootte:** →

$$E = \frac{\bar{X}_1 - \bar{X}_2}{\frac{1}{2}(S_1 + S_2)}$$

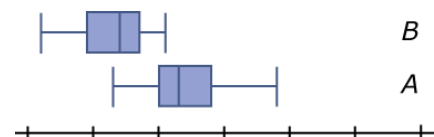
(Let op: boven de streep altijd grootste waarde - kleinste waarde, dus uitkomst altijd positief)

De volgende conclusies kunnen nu worden getrokken:

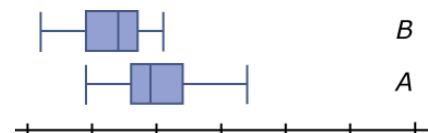
- als  $E > 0,8$ , dan zeggen we “*het verschil is groot*”;
- als  $0,4 < E \leq 0,8$ , dan zeggen we “*het verschil is middelmatig*”;
- als  $E \leq 0,4$ , dan zeggen we “*het verschil is gering*”.

Tenslotte kun je ook de **boxplots** (zie hoofdstuk 2) van de twee groepen met elkaar vergelijken. Daarbij gelden de volgende regels:

- als de boxen<sup>1)</sup> elkaar niet overlappen, dan zeggen we “*het verschil is groot*”,



- als de boxen elkaar wel overlappen en een mediaan van een boxplot buiten de box van de andere boxplot ligt, dan zeggen we “*het verschil is middelmatig*”,



- in alle andere gevallen, waarbij de boxplots elkaar nog meer overlappen, zeggen we “*het verschil is gering*”.

1) de box is het “rode, rechthoekige deel” van Q1 tot Q3.  $\delta$