

FORMULEBLAD STATISTIEK

Vuistregels voor de grootte van het verschil van twee groepen

2 x 2 kruistabel $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ met $\phi = \varphi = \frac{ad-bc}{\sqrt{(a+b)(a+c)(b+c)(b+d)}}$

- als $\varphi < -0,4$ of $\varphi > 0,4$, dan zeggen we “het verschil is groot”,
- als $-0,4 \leq \varphi < 0,2$ of $0,2 < \varphi \leq 0,4$, dan zeggen we “het verschil is middelmatig”,
- als $-0,2 \leq \varphi \leq 0,2$, dan zeggen we “het verschil is gering”.

Maximaal verschil in cumulatief percentage ($\max(V_{cp})$) met steekproefomvang $n > 100$

- als $\max(V_{cp}) > 40$, dan zeggen we “het verschil is groot”,
- als $20 < \max(V_{cp}) \leq 40$, dan zeggen we “het verschil is middelmatig”,
- als $\max(V_{cp}) \leq 20$, dan zeggen we “het verschil is gering”.

Effectgrootte $E = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{1}{2}(S_1^2 + S_2^2)}}$, met $\bar{X}_1$ en $\bar{X}_2$ de steekproefgemiddelden ($\bar{X}_1 > \bar{X}_2$), S_1 en S_2 de steekproefstandaardafwijkingen

- 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”.

Twee **boxplots** vergelijken

- als de boxen ¹⁾ 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 zeggen we “het verschil is gering”

Betrouwbaarheidsintervallen

Het 95%-betrouwbaarheidsinterval voor de **populatieproportie** is

$$p \pm 1,96 \cdot \sqrt{\frac{p(1-p)}{n}}, \text{ met } p \text{ de steekproefproportie en } n \text{ de steekproefomvang.}$$

Het 95%-betrouwbaarheidsinterval voor het **populatiegemiddelde** is

$$\bar{X} \pm 1,96 \cdot \frac{S}{\sqrt{n}}, \text{ met } \bar{X} \text{ de steekproefproportie, } n \text{ de steekproefomvang en } S \text{ de steekproefstandaardafwijking.}$$

¹⁾ De ‘box’ is het interval vanaf het eerste kwartiel tot en met het derde kwartiel.

Vuistregels voor de correlatie tussen twee kwantitatieve variabelen

Voor de **correlatiecoëfficiënt** r geldt

- Als $r \leq -0,7$ dan is er sprake van sterke negatieve samenhang.
- Als $-0,7 < r \leq -0,3$ dan is er sprake van matige negatieve samenhang.
- Als $-0,3 < r < 0$ dan is er sprake van zwakke negatieve samenhang.
- Als $0 < r < 0,3$ dan is er sprake van zwakke positieve samenhang.
- Als $0,3 \leq r < 0,7$ dan is er sprake van matige positieve samenhang.
- Als $r \geq 0,7$ dan is er sprake van sterke positieve samenhang.