

Regressielijnen

In een **spreadsdiagram** van de variabelen x en y zet je alle combinaties (x,y) als een **puntenwolk** in een assenstelsel. Of er een **statistisch verband** bestaat tussen x en y wordt bepaald door de **correlatiecoëfficiënt** r_{xy} .

- $r_{xy} = 1$: perfecte positieve correlatie tussen x en y ; de punten liggen op een stijgende lijn.
- $r_{xy} = 0$: geen enkele correlatie tussen x en y .
- $r_{xy} = -1$: perfecte negatieve correlatie tussen x en y ; de punten liggen op een dalende lijn.

De correlatie wordt beter naarmate r_{xy} dichter bij 1 of -1 ligt.

Gebruik de **r-tabel**. In Excel wordt vaak de determinatiecoëfficiënt r_{xy}^2 gegeven.

Een verband waarbij de toename (of afname) van de éne variabele een gevolg is van een toename (of afname) van de andere heet een **causaal verband**: er is dan sprake van oorzaak en gevolg. Een **statistisch verband** tussen twee variabelen hoeft niet causaal te zijn.

Bij voldoende correlatie kun je een formule van de vorm $y = ax + b$ opstellen, de **regressielijn** van y op x . Zo'n regressielijn gaat door het punt (\bar{x}, \bar{y}) en heeft als

richtingscoëfficiënt (hellingsgetal): $a = r_{xy} \cdot \frac{s_y}{s_x}$

