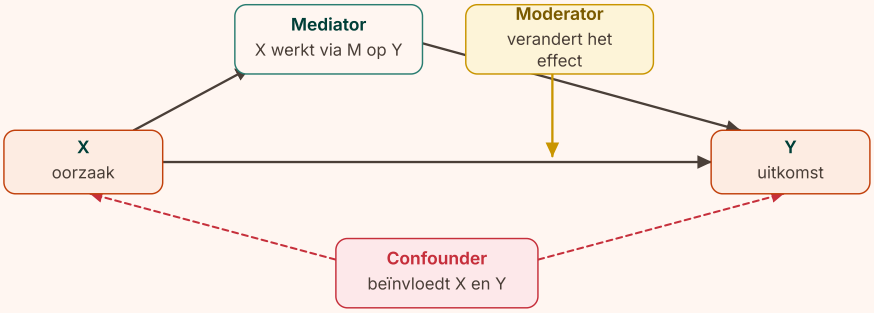
Diagnostische analytics en causaliteit

CHEATSHEET

Tentamen ma 7 dec 2026

TENTAMENKANS Mediator en moderator **94%** · Correlatie is geen causaliteit **93%** · Causaal relatieschema **92%** · Toetsbare hypothese en nulhypothese formuleren **92%** · Meetniveaus: nominaal, ordinaal, continu, discreet **92%** · Hypothesen vooraf formuleren (data dredging) **90%**

Causaal relatieschema



Soorten analytics

1. **Descriptive: wat gebeurde er?**
2. **Diagnostic: waarom gebeurde het?**
3. **Predictive: wat gebeurt er straks?**
4. **Prescriptive: wat moeten we doen?**

M 01 Formules (formuleblad of kennen)

- SE = s / wortel(n) (met sigma bij bekende sigma)
- CI = x-streep +/- t* x s / wortel(n) (formuleblad); of z* x sigma / wortel(n) bij bekende sigma of grote n (formuleblad)
- z* : 90% = 1,64 | 95% = 1,96 | 99% = 2,58 (kennen)
- Z = (x – mu) / sigma (kennen)
- Normale verdeling: 68% binnen 1 sd, 95% binnen 2 sd, 99,7% binnen 3 sd (kennen)
- Variatiecoëfficiënt = sd / mean; IQR = Q3 – Q1; range = max – min

M 02 R-functies

- mean(), median(), sd(), var(), range(), quantile(x, p), IQR(), summary(), table(), prop.table()
- pnorm(x, mu, sigma) = kans links van x; rechterstaart = 1 – pnorm(). qnorm(p) = kritieke waarde. qt(p, df)
- cor(x, y): samenhang, nooit bewijs van oorzaak

M 03 Lezen van summary()

- Mean > Median: rechts-scheef. Mean < Median: links-scheef. 1st Qu. en 3rd Qu. geven de IQR.

M 04 Beslisregels

- SE kleiner als n groter of s kleiner. SE halveren: 4 x zoveel n.
- Hoger betrouwbaarheidsniveau geeft breder CI.
- 95% CI: "bij herhaling bevat 95% van de intervallen mu". Nooit: "95% kans dat mu in dit interval zit" of "95% van de data".
- n groter dan 20 tot 30: CLT, gemiddelde bij benadering normaal.

M 05 Causal relationship scheme

- Direct effect X naar Y. **Mediator**: X naar M naar Y (verklaart het kanaal). **Moderator**: Z verandert de sterkte van X naar Y (interactie). **Confounder**: Z veroorzaakt X en Y (schijnverband). **Reverse causality**: Y beïnvloedt ook X.
- Symbolen: + positief, – negatief, * categorisch, ? concurrerende hypothesen.

M 06 Hypothesen

- Toetsbaar, richting, variabelen helder, H0 = tegenhanger (geen effect of geen positieve invloed).
- Associatief (enquête) of causaal (experiment).
- Vooraf, tegen data dredging en confirmation bias; sturen de dataverzameling.

M 07 Variabelen en validiteit

- Nominaal, ordinaal (kwalitatief); continu, discreet (kwantitatief). Afhankelijk = uitkomst Y, onafhankelijk = verklaring X.
- Internal validity: betrouwbaar voor de bronpopulatie. External validity: generaliseren naar target population of andere tijden.