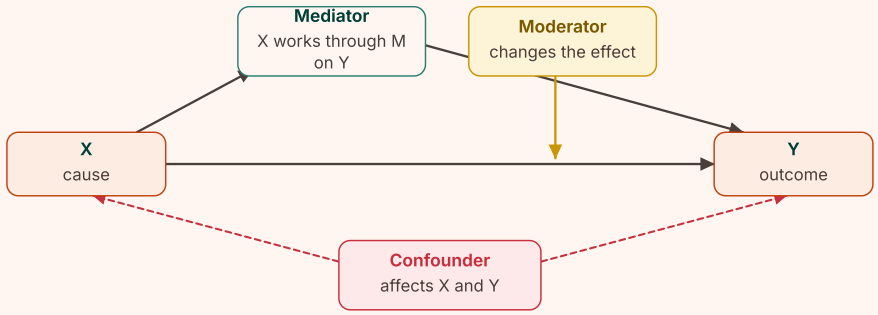
Diagnostic analytics and causality

CHEATSHEET

Exam Mon, 7 Dec 2026

EXAM LIKELIHOOD Mediator and moderator **94%** · Correlation is not causation **93%** · Causal relationship scheme **92%** · Formulating a testable hypothesis and null hypothesis **92%** · Measurement levels: nominal, ordinal, continuous, discrete **92%** · Formulating hypotheses before data collection (data dredging) **90%**

Causal relationship diagram



Types of analytics

1. Descriptive: what happened?
2. Diagnostic: why did it happen?
3. Predictive: what will happen?
4. Prescriptive: what should we do?

M 01 Formulas (formula sheet or know them)

- SE = s / sqrt(n) (use sigma if known)
- CI = x-bar +/- t* x s / sqrt(n) (formula sheet); or z* x sigma / sqrt(n) for known sigma or large n (formula sheet)
- z* : 90% = 1.64 | 95% = 1.96 | 99% = 2.58 (know them)
- Z = (x – mu) / sigma (know it)
- Normal distribution: 68% within 1 sd, 95% within 2 sd, 99.7% within 3 sd (know it)
- Coefficient of variation = sd / mean; IQR = Q3 – Q1; range = max – min

M 02 R functions

- mean(), median(), sd(), var(), range(), quantile(x, p), IQR(), summary(), table(), prop.table()
- pnorm(x, mu, sigma) = probability left of x; right tail = 1 – pnorm(). qnorm(p) = critical value. qt(p, df)
- cor(x, y): association, never proof of cause

M 03 Reading summary()

- Mean > Median: right-skewed. Mean < Median: left-skewed. 1st Qu. and 3rd Qu. give the IQR.

M 04 Decision rules

- SE smaller if n larger or s smaller. To halve SE: 4 x as many n.
- Higher confidence level gives a wider CI.
- 95% CI: "on repetition 95% of the intervals contain mu". Never: "95% chance that mu is in this interval" or "95% of the data".
- n above 20 to 30: CLT, mean approximately normal.

M 05 Causal relationship scheme

- Direct effect X to Y. **Mediator**: X to M to Y (explains the channel). **Moderator**: Z changes the strength of X to Y (interaction). **Confounder**: Z causes X and Y (spurious relationship). **Reverse causality**: Y also influences X.
- Symbols: + positive, – negative, * categorical, ? competing hypotheses.

M 06 Hypotheses

- Testable, direction, variables clear, H0 = counterpart (no effect or no positive influence).
- Associative (survey) or causal (experiment).
- In advance, against data dredging and confirmation bias; they steer data collection.

M 07 Variables and validity

- Nominal, ordinal (qualitative); continuous, discrete (quantitative). Dependent = outcome Y, independent = explanation X.
- Internal validity: trustworthy for the source population. External validity: generalise to the target population or other times.