Start:Bayes!

Die Atome der Kausalität

Sebastian Sauer



Lernziele

Nach diesem Kapitel können Sie …

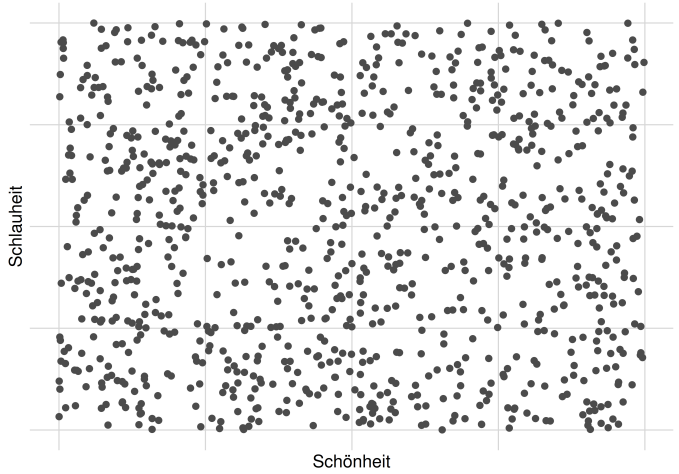
- erklären, wann eine Kausalaussage gegeben eines DAGs berechtigt ist
- erklären, warum ein statistisches Modell ohne Kausalmodell zumeist keine Kausalaussagen treffen kann
- die “Atome” der Kausalität eines DAGs benennen
- “kausale Hintertüren” schließen

Begleitliteratur: ähnliche Darstellungen finden sich bei Rohrer (2018).

Kollision

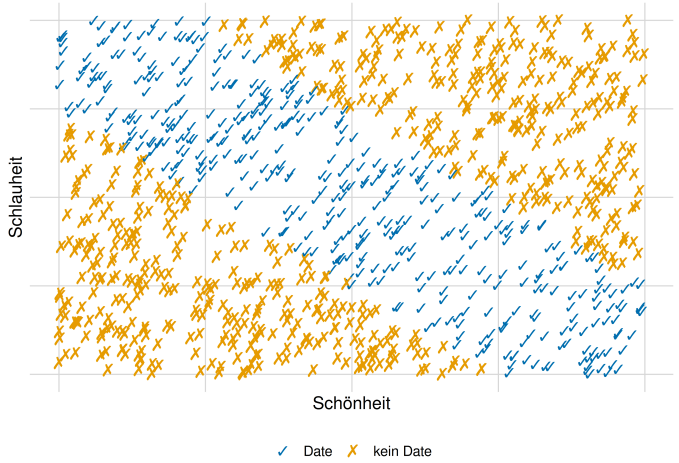


Kein Zusammenhang von Intelligenz und Schönheit (?)



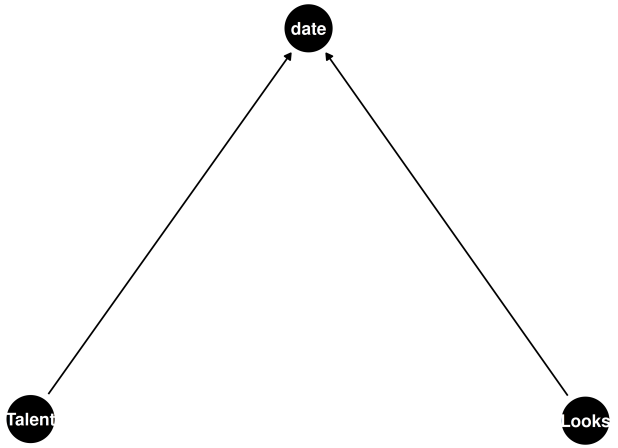
- In den Daten: kein Zusammenhang von Intelligenz (`talent`) und Schönheit (`looks`)
- Für geringe wie hohe Intelligenzwerte: breites Spektrum an Schönheitswerten

Aber Ihre Dates sind entweder schlau oder schön



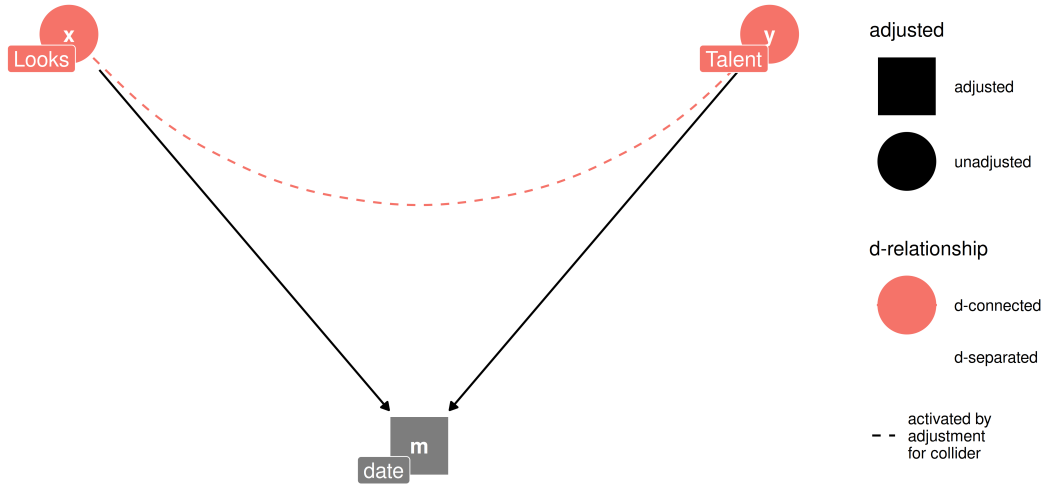
- Die Menschen, die Sie daten, sind entweder schön **oder** schlau — selten beides
- Wie kann das sein, wenn Schönheit und Intelligenz unabhängig sind?

DAG zur Rettung



- `date` ist eine **gemeinsame Wirkung** von `Looks` und `Talent`
- Stratifiziert man nach `date` (z. B. “nur Dates betrachten”), entsteht eine **Scheinkorrelation** zwischen Schönheit und Intelligenz

Kollision — zwei Darstellungen



Kontrolliert man die gemeinsame Wirkung m, entsteht eine Scheinkorrelation zwischen den beiden Ursachen x und y.

Definition: Kollision

Definition: Kollision

Als **Kollision** (Kollisionsverzerrung, Auswahlverzerrung, engl. *collider*) bezeichnet man einen DAG, bei dem eine Wirkung zwei Ursachen hat (eine gemeinsame Wirkung zweier Ursachen) (Pearl et al., 2016, p. 40).

Wichtig

Zu viele oder falsche Prädiktoren in einer Regression können bewirken, dass die Koeffizienten nicht die kausalen Effekte zeigen, sondern durch Scheinkorrelation verzerrte Werte.

Tipp

🤦‍♀️ Kontrollieren Sie keine Kollisionsvariablen.

Einfaches Beispiel zur Kollision

In der Zeitschrift *Glitzer* werden nur gezeigt:

- Schöne Menschen 🪞
- Reiche Menschen 🤑

Schönheit und Reichtum seien unabhängig voneinander.

Frage: Don ist nicht schön, taucht aber häufig in der Glitzer auf — was lernen wir über seine finanzielle Situation?

→ Don muss reich sein.

Formal: Kollision als bedingte Abhängigkeit

Sei $Z = X + Y$, wobei X und Y unabhängig sind.

- Wissen Sie nur $X = 3$: Sie lernen nichts über Y
- Wissen Sie zusätzlich $Z = 10$: Sie wissen sofort $Y = 7$

Also sind X und Y abhängig, **gegeben** Z:

$$X \not\perp\!\!\!\perp Y \mid Z$$



Kontrollieren erzeugt die Kollisionsverzerrung

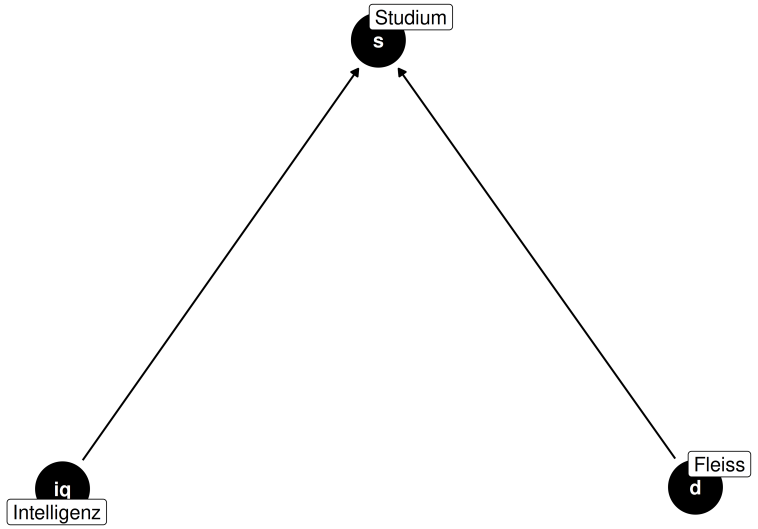
- *Stratifizieren*: `date` in Gruppen aufteilen, dann `talent`–`looks` je Gruppe betrachten
- *Kontrollieren mit Regression*: `date` als zusätzlicher Prädiktor neben `looks` bzw. `talent`

Ohne Kontrolle von `date`: **keine** Scheinkorrelation — der Pfad ist blockiert.

Mit Kontrolle von `date`: der Pfad `Looks → date → talent` öffnet sich — Scheinkorrelation entsteht.

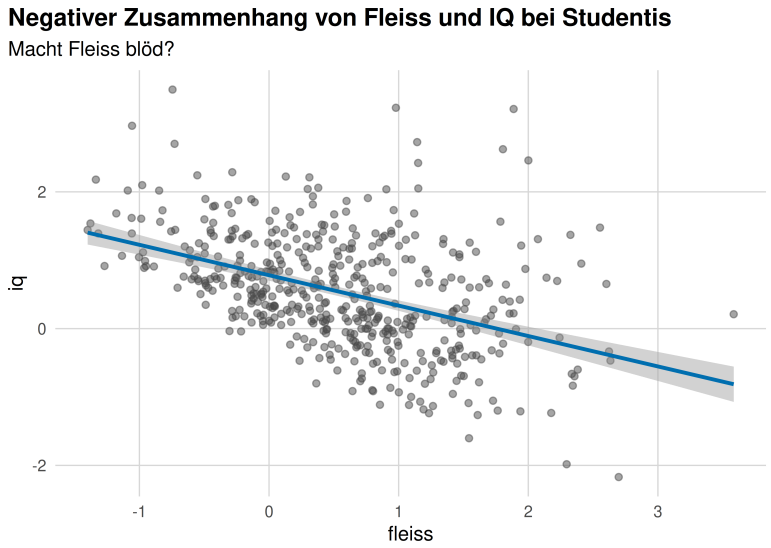


IQ, Fleiß und Eignung fürs Studium



- `Eignung` = Summe aus Intelligenz (`iq`) und Fleiß (`fleiss`), plus etwas Glück
- Nur wer geeignet ist, studiert (`studium = 1`)

Schlagzeile “Fleiß macht blöd!”

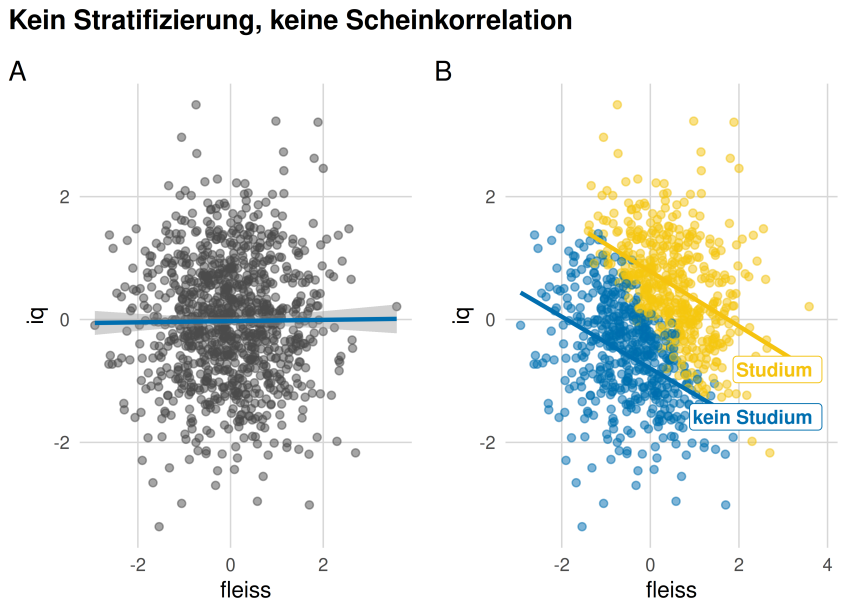


- Analyse **nur** unter Studierenden (`studium == 1`): negativer Zusammenhang von Fleiß und IQ
- Ursache: Kollisionsverzerrung durch Selektion (nur wer studiert, wird betrachtet)

Definition: Stratifizieren

Definition: Stratifizieren

Durch **Stratifizieren** wird eine Stichprobe in (homogene) Untergruppen unterteilt (sog. *Strata*).



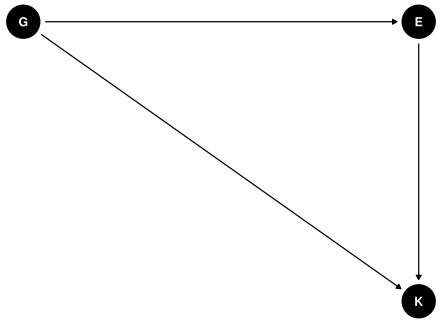
Ohne Stratifizierung: keine Scheinkorrelation. **Mit** Stratifizierung: Scheinkorrelation.

Kontrollieren kann schaden oder nützen

- Wildes Kontrollieren einer Variablen kann genauso gut schaden wie nützen
- **Nur** Kenntnis des DAGs verrät die richtige Entscheidung

> **Hinweis**
>
> Nimmt man eine Variable als zweiten Prädiktor auf, “kontrolliert” man diese Variable — das Regressionsgewicht des ersten Prädiktors wird um den Einfluss des zweiten “bereinigt”.

Einfluss von Großeltern und Eltern auf Kinder

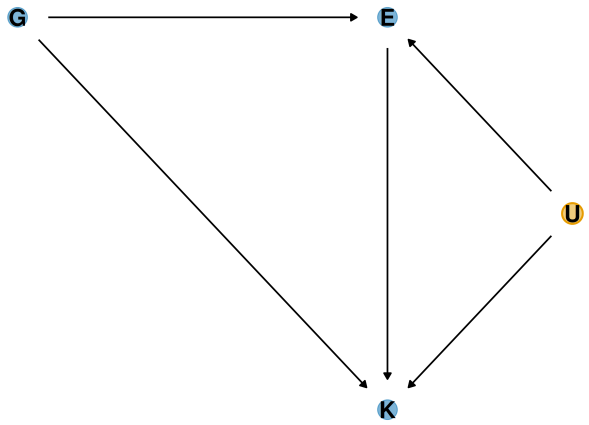


Angenommene Effekte:

- indirekt: $G \to E \to K$
- direkt: $E \to K$
- direkt: $G \to K$

Interessiert: der direkte kausale Effekt $G \to K$.

Der Gespenster-DAG 👻



- Es gibt “unheilbare” DAGs, in denen kein unverzerrter Kausaleffekt bestimmbar ist
- U: unbeobachteter Einfluss (z. B. Nachbarschaft) auf E und K
- E ist Kollisionsvariable auf diesem Pfad → Kontrolle von E verzerrt $G \rightarrow K$, selbst ohne U je zu messen

Fazit: Diesen DAG müssen wir verloren geben — zurück an den Schreibtisch, oder mehr Daten sammeln.

Die Hintertür schließen

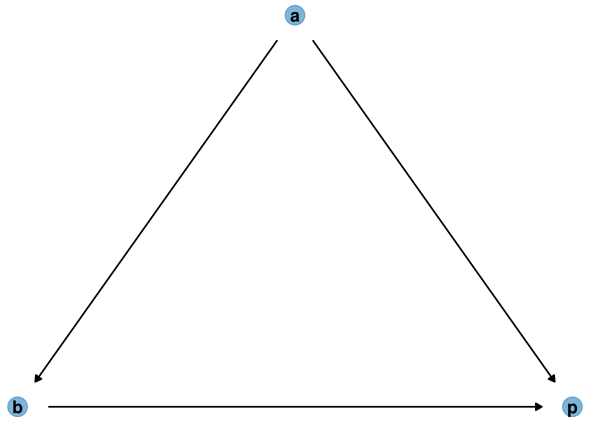
Definition: Hintertür

Definition: Hintertür

Eine **Hintertür** ist ein nicht-kausaler Pfad zwischen einer UV und einer AV. Ein Hintertürpfad entsteht, wenn es eine alternative Route über eine oder mehrere Variablen gibt, die UV mit AV verbindet. Dieser Pfad verzerrt die Schätzwerte des kausalen Einflusses, wenn er nicht kontrolliert wird.

Zur Erinnerung: Konfundierung

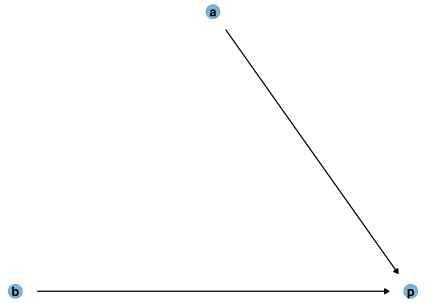
Forschungsfrage: Wie groß ist der kausale Einfluss der Schlafzimmerzahl (b) auf den Verkaufspreis (p) eines Hauses?



Zwei Pfade zwischen b und p:

1. $b \rightarrow p$ (kausal)
2. $b \leftarrow a \rightarrow p$ (nicht-kausal, über Wohnfläche a)

Gute Experimente zeigen den echten kausalen Effekt



- Ist die Hintertür geschlossen (keine Pfeile führen in die UV), ist eine Konfundierung ausgeschlossen
- Ein randomisiertes Experiment entfernt den Pfeil $a \rightarrow b$ und schließt damit den zweiten Pfad

Wichtig

Gut gemachte Experimente schließen kausale Hintertüren — das erlaubt korrekte Kausalaussagen.

Hintertür schließen auch ohne Experimente

- Konfundierende Pfade blockieren = “die Hintertür schließen” (*backdoor criterion*)
- Z. B. die Konfundierungsvariable a als weiteren Prädiktor in die Regression aufnehmen

> **Tipp**
>
> Kontrollieren Sie Konfundierer, um kausale Hintertüren zu schließen.

$$b \perp\!\!\!\perp p \mid a$$

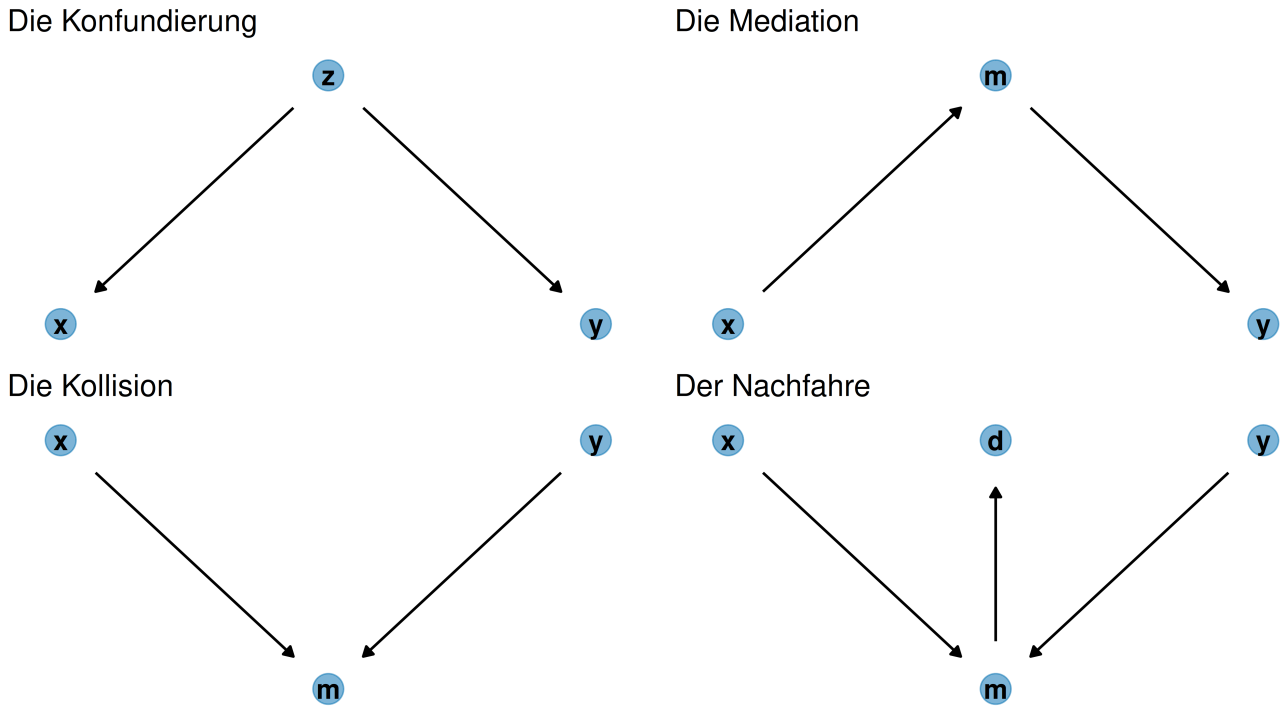
Die Hintertür schließen



Die vier Atome der Kausalanalyse



Die vier Grundbausteine im Überblick



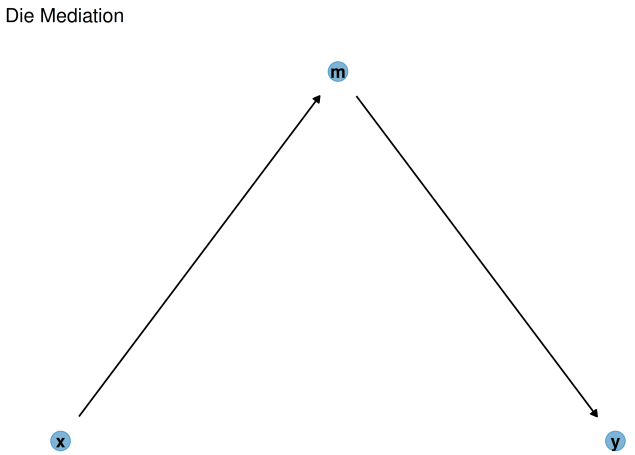
Kennen Sie diese vier Grundbausteine, können Sie jedes beliebige Kausalsystem (DAG) entschlüsseln:

Konfundierung, Mediation, Kollision, Nachfahre

Definition: Mediator

Definition: Mediator

Einen Pfad mit drei Variablen, verbunden über zwei Kanten, wobei die Pfeile von UV zu Mediator und von Mediator zur AV zeigen, nennt man **Mediation**. Der **Mediator** ist die Variable zwischen UV und AV (Pearl et al., 2016, p. 38).



$$x \rightarrow m \rightarrow y$$

Mediator kontrollieren?

Frage: Sollte man den Mediator m kontrollieren, wenn man den Kausaleffekt von x auf y schätzen möchte?

Nein — durch Kontrollieren von m wird der Kausalpfad von x zu y geschlossen; es kann keine kausale Assoziation mehr "fließen".

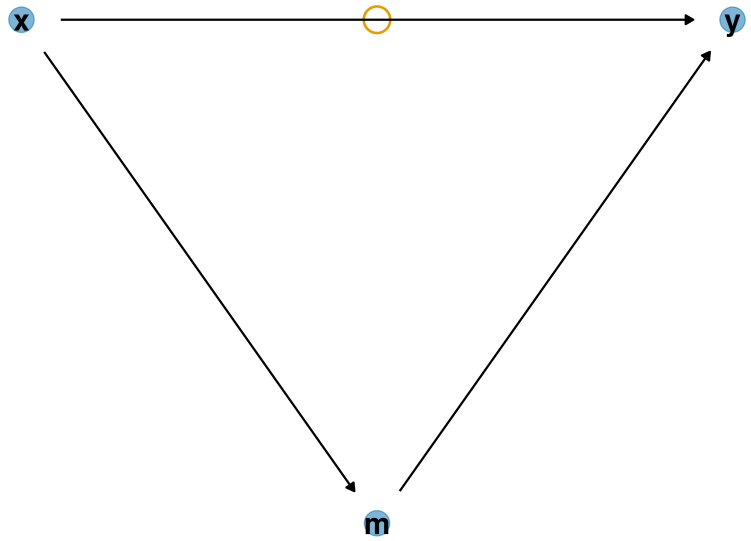
Tipp

Kontrollieren Sie den Mediator *nicht*. Der Pfad über den Mediator ist ein "echter" Kausalpfad, keine Scheinkorrelation.

Wichtig

Das Kontrollieren eines Mediators ist ein Fehler, wenn man am gesamten (totalen) Kausaleffekt von UV zu AV interessiert ist.

Partielle Mediation



Zwei kausale Pfade von X zu Y: $x \rightarrow m \rightarrow y$ und $x \rightarrow y$

Effekt-Begriffe

Definition: Kausaleffekt

Gibt es eine (von praktisch Null verschiedene) kausale Assoziation der UV auf die AV, spricht man von einem Kausaleffekt.

Definition: Totaler Effekt

Die Summe der Effekte aller kausalen Pfade von UV zu AV.

Definition: Indirekter Effekt

Der kausale Effekt über den Mediatorpfad (von X über M zu Y).

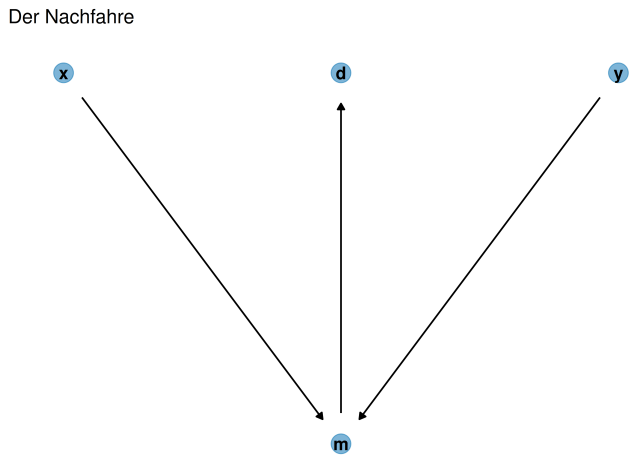
Definition: Direkter Effekt

Ein Effekt, der nur aus dem Pfad $x \rightarrow y$ besteht, ohne Zwischenglieder.

Definition: Nachfahre

Definition: Nachfahre

Ein **Nachfahre** (engl. *descendant*) ist eine Variable, die von einer anderen Variable (kausal) beeinflusst wird.



Kontrolliert man einen Nachfahren d, kontrolliert man teilweise auch dessen Vorfahren m — der Pfad von x nach y öffnet sich teilweise, da m eine Kollisionsvariable ist.

Kochrezept zur Analyse von DAGs 👨🍳

Egal wie kompliziert ein DAG aussieht — er besteht immer aus den vier Atomen.

1. Listen Sie alle Pfade von UV (X) zu AV (Y) auf
2. Beurteilen Sie für jeden Pfad: geschlossen oder geöffnet?
3. Beurteilen Sie für jeden Pfad: Hintertürpfad? (Pfeil führt in die UV)
4. Bei offenen Hintertürpfaden: welche Variablen kontrollieren, um sie zu schließen?

Schließen Sie die Hintertür (wenn möglich)!

Hintertür: ja oder nein?

Fall 1: $\boxed{X \rightarrow}$

Pfade, die von der UV **wegführen**: entweder “gute” Kausalpfade oder automatisch geblockte Nicht-Kausal-Pfade → nichts zu tun (Nachkommen der UV nicht kontrollieren!)

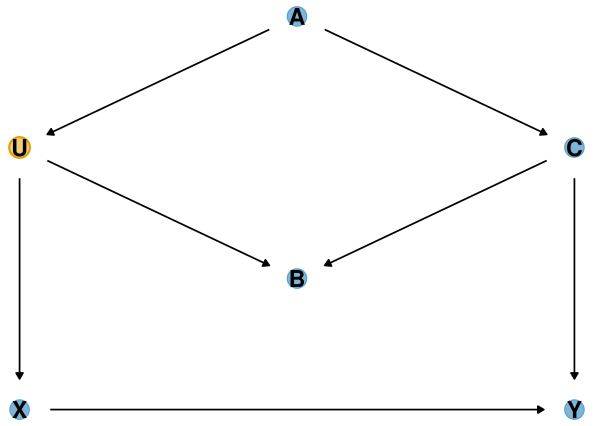
Fall 2: $\boxed{\rightarrow X}$

Pfade, die zur UV **hinführen**: immer Nicht-Kausal-Pfade (Hintertüren) → offene Hintertüren schließen

> **Tipp**
>
> Schließen Sie immer offene Hintertüren, um Verzerrungen der Kausaleffekte zu verhindern.

Beispiel: `bsp1`

UV: X, AV: Y, Kovariaten A, B, C, unbeobachtete Variable U



Zwei Hintertürpfade:

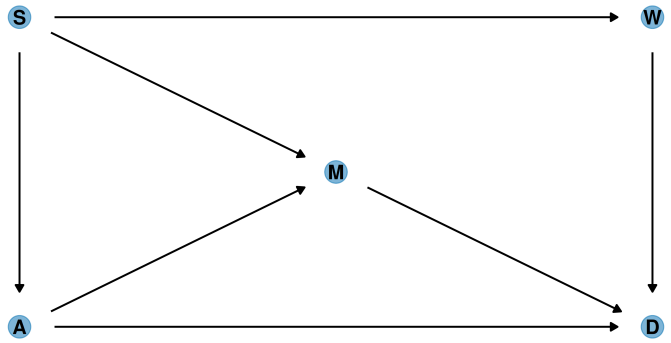
1. $X \leftarrow U \leftarrow A \rightarrow C \rightarrow Y$ — **offen**
2. $X \leftarrow U \rightarrow B \leftarrow C \rightarrow Y$ — geschlossen

→ Kontrollieren von A oder C schließt die offene Hintertür.

Schließen Sie die Hintertür (wenn möglich)!

Beispiel: bsp2

UV: W, AV: D



Um offene Hintertüren zu schließen, kontrollieren Sie:

- entweder A und M
- oder S

Implizierte bedingte Unabhängigkeiten

- Auch wenn Daten nicht sagen können, welcher DAG richtig ist — wir können lernen, welcher DAG **falsch** ist
- Ein DAG impliziert bestimmte bedingte Unabhängigkeiten zwischen Variablen
- Diese können mit den Daten geprüft werden → Möglichkeit, einen DAG zu verwerfen



Fazit



Zusammenfassung

Interpretierbarkeit hängt von der epistemologischen Zielrichtung der Forschungsfrage ab:

- **Deskriptiv**: Ergebnisse direkt interpretierbar, aber keine kausale Interpretation zulässig
- **Prognostisch**: Modellkoeffizienten spielen keine Rolle, es geht um vorhergesagte Werte $\hat{y}_i$; Kausalaussagen nicht möglich, aber auch nicht von Interesse
- **Kausal**: Modellkoeffizienten nur auf Basis eines Kausalmodells (DAG) oder eines guten Experiments interpretierbar

Möglichst viele Prädiktoren ins Modell aufzunehmen ist **falsch**, wenn das Ziel eine Kausalaussage ist. Kenntnis der “kausalen Atome” ist Voraussetzung für Kausalschlüsse in Beobachtungsstudien.



Vielen Dank!

Fragen?

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley.

