Anwendungsorientierte Analyseverfahren

Regressionsannahmen

Prof. Dr. Michael Scharkow

Sommersemester 2024

ANNAHMEN DES GLM

Statistische Annahmen

- Linearität und Additivität der Zusammenhänge
- Normalverteilung und Homoskedastizität der Residuen
- Unabhängigkeit der Residuen
- keine einflussreichen Ausreißer
- keine Multikollinearität

Kausalannahmen

- korrekt spezifiziertes Modell, d.h. keine fehlenden oder überflüssigen Kovariaten

LINEARITÄT & ADDITIVITÄT

- **Annahme**: der Zusammenhang zwischen X und Y ist linear und unabhängig von Z
- **Diagnose**: Inspektion des Scatterplots bzw. des Fitted/Residual-Plots
- **Verletzung**: nichtlineare Zusammenhänge (quadratisch, exponentiell, etc.)
- **Konsequenz der Verletzung**: verzerrte Regressionskoeffizienten
- **Lösung**: Transformation von X oder Y, nichtlineares Regressionsmodell, Moderationsanalyse mit Z

HOMOSKEDASTIZIZÄT DER RESIDUEN

- **Annahme**: Residualvarianz ist für alle Werte von X gleich
- **Diagnose**: Fitted/Residual-Plots
- **Verletzung**: Residuen streuen in Abhängigkeit von X
- **Konsequenz der Verletzung**: falsche Standardfehler, ineffiziente Schätzung
- **Lösung**: alternative Standardfehler, Datentransformationen, alternatives Modell

UNABHÄNGIGKEIT DER RESIDUEN

- **Annahme**: Residuen korrelieren weder miteinander noch mit den Prädiktoren
- **Diagnose**: Nachdenken über datengenerierenden Prozess, Test auf serielle Korrelation
- **Verletzung**: Residuen (und oft Variablen) sind geclustert (zeitlich, Stichprobe)
- **Konsequenz der Verletzung**: falsche Standardfehler, ineffiziente Schätzung
- **Lösung**: Multilevel-Modell, Modell mit Autokorrelationen

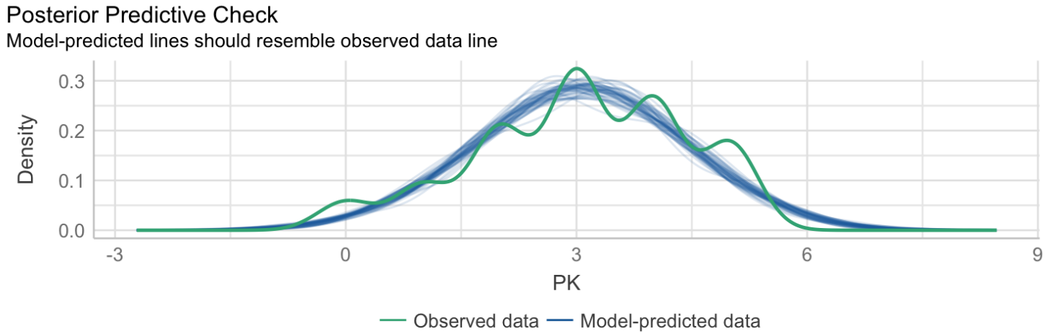
KEINE EINFLUSSREICHEN AUSREISSER

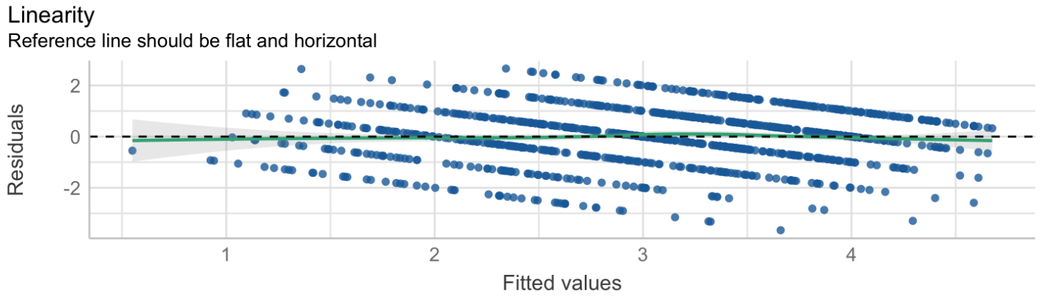
- **Annahme**: alle Fälle tragen gleich zur Schätzung bei
- **Diagnose**: Scatterplot, Leverage-Plot
- **Verletzung**: einzelne Fälle beeinflussen die Höhe der Regressionsgeraden
- **Konsequenz der Verletzung**: verzerrte Regressionskoeffizienten
- **Lösung**: Ausschluss von Ausreißern (mit klar definierten Regeln!)

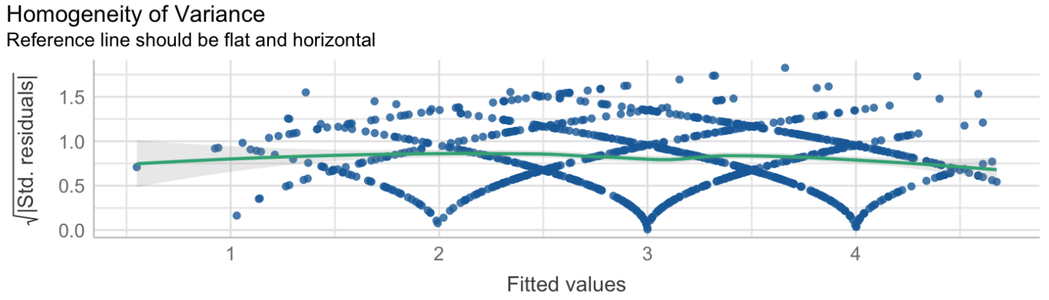
KEINE MULTIKOLLINEARITÄT

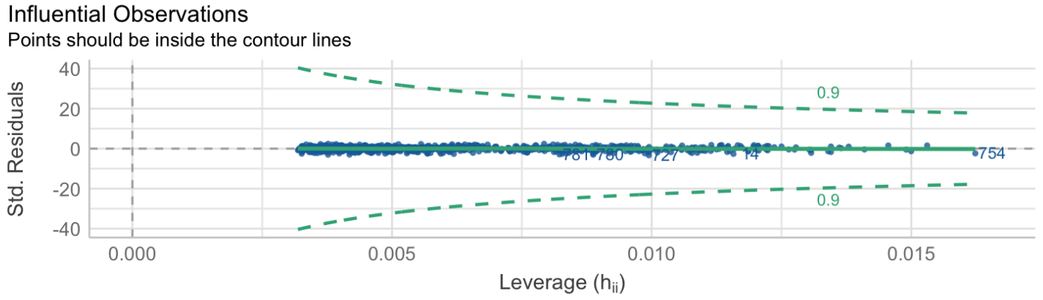
- **Annahme**: Prädiktorvariablen X korrelieren nicht zu stark miteinander
- **Diagnose**: Korrelationsmatrix der Prädiktoren, VIF-Analyse (Variance Inflation Factor)
- **Verletzung**: Prädiktorvariablen korrelieren stark miteinander
- **Konsequenz der Verletzung**: falsche Standardfehler, ineffiziente Schätzung
- **Lösung**: Ausschluss von Prädiktorvariablen

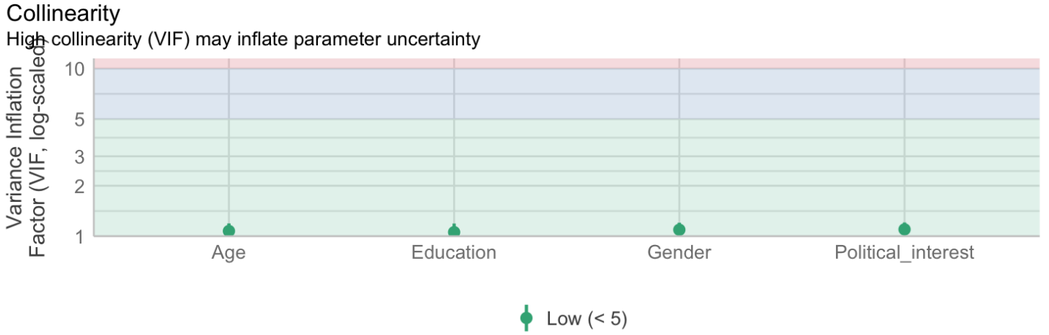
BEISPIEL: VAN ERKEL & VAN AELST, 2021

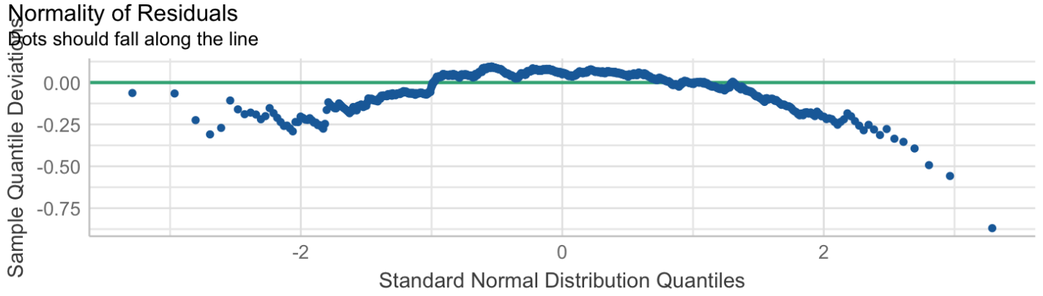




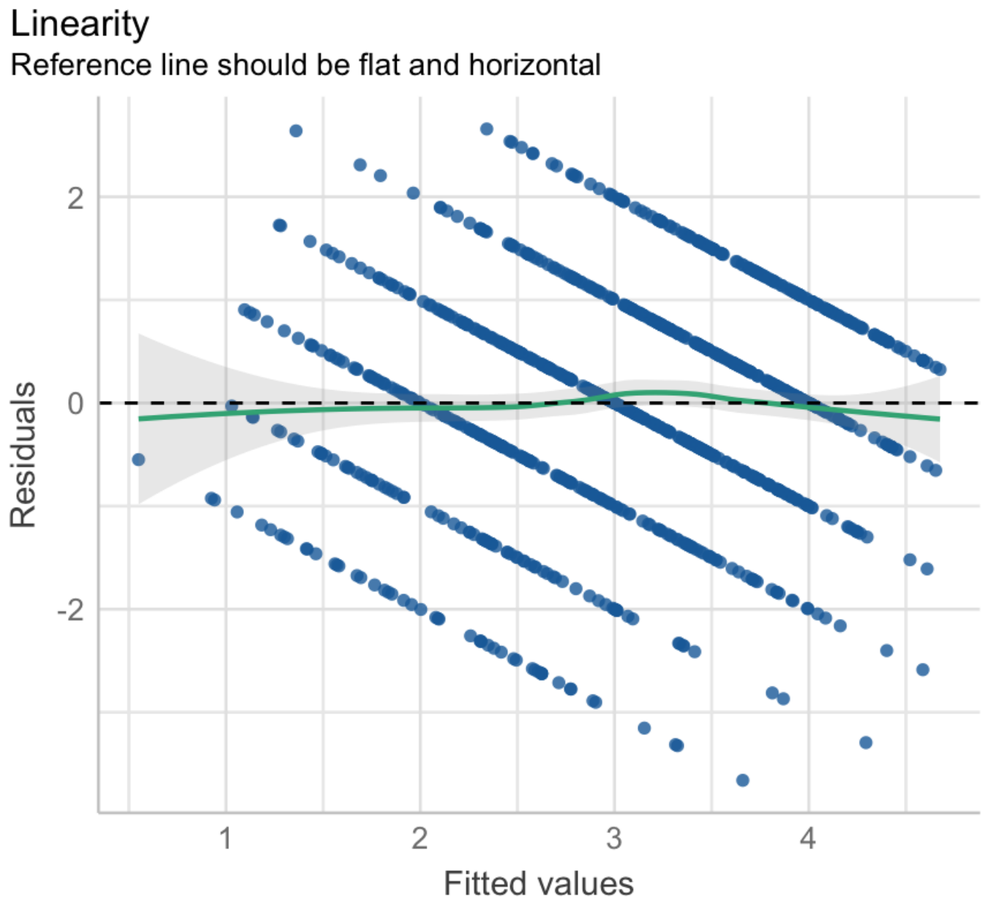


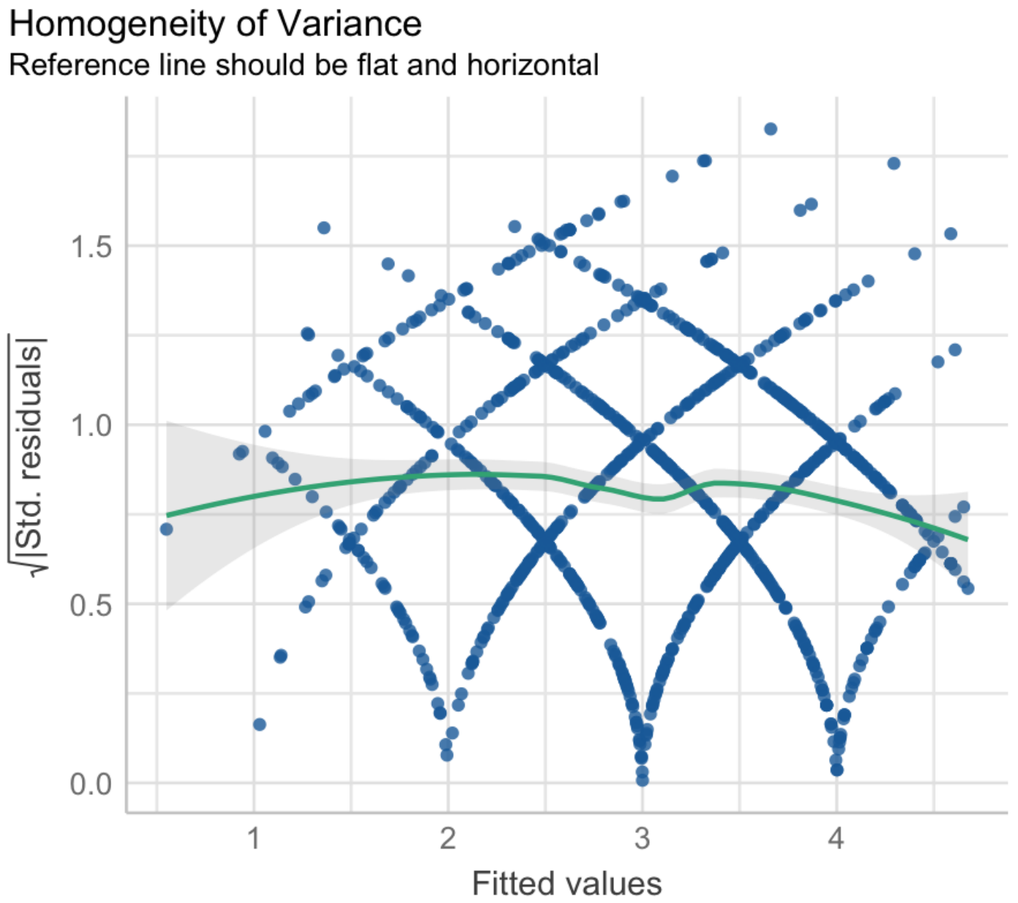






LINEARITÄT UND HOMOSKEDASTIZITÄT

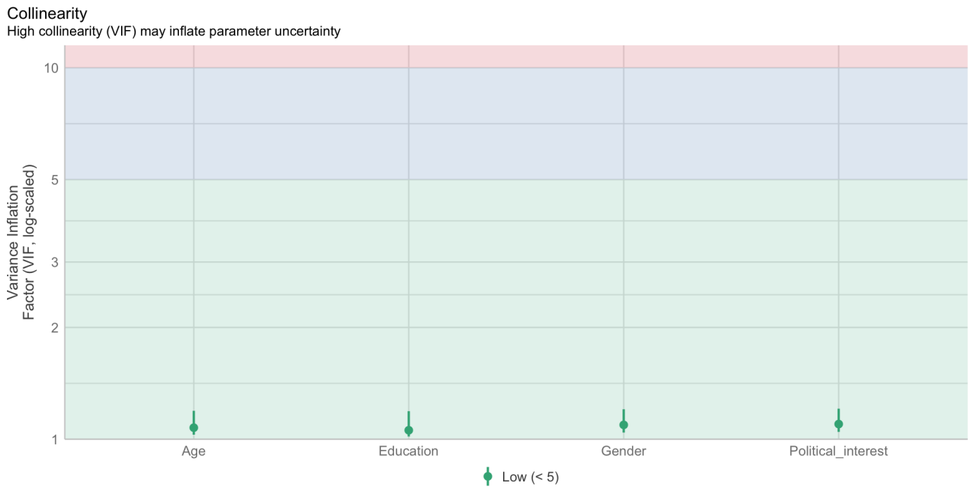




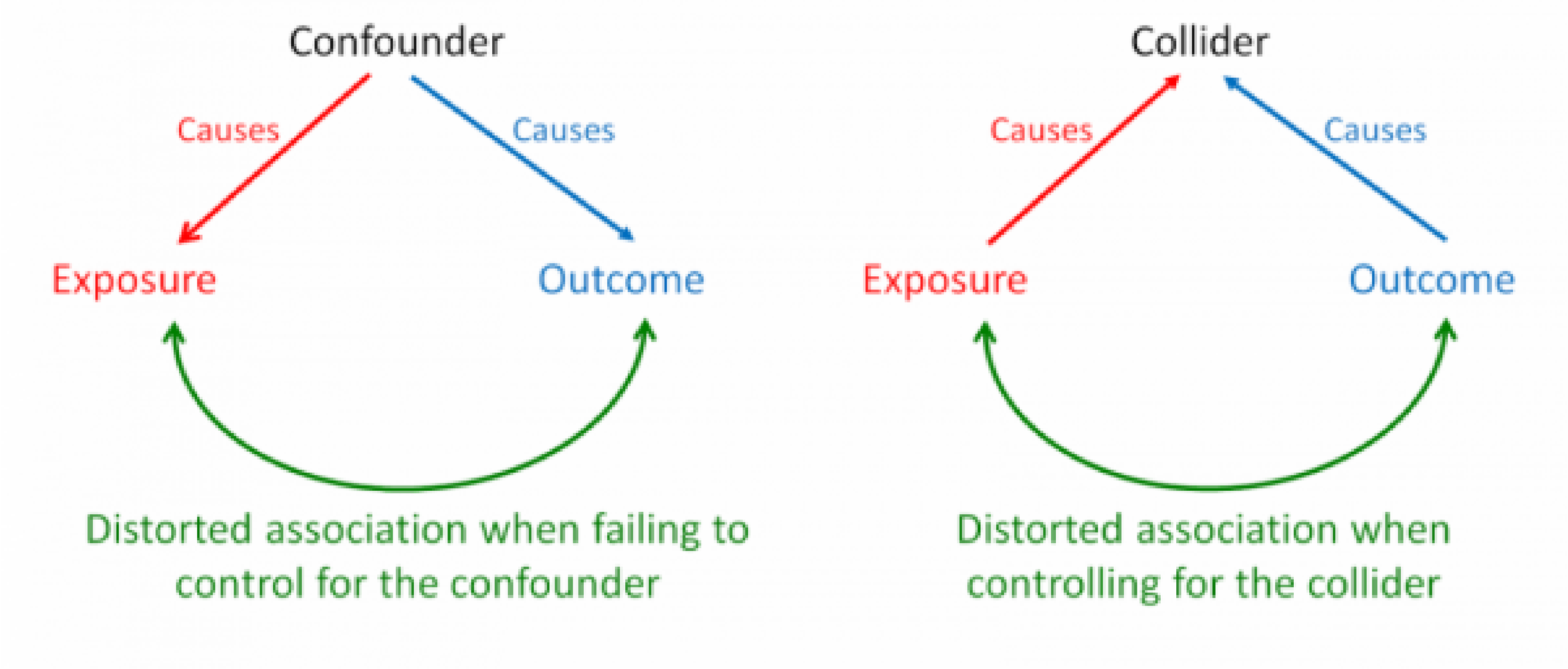
MULTIKOLLINEARITÄT

Parameter	Political_interest	Age	Gender
PK	0.49	0.3	-0.31
Gender	-0.21	-0.2	NA
Age	0.14	NA	NA

```
$VIF
```

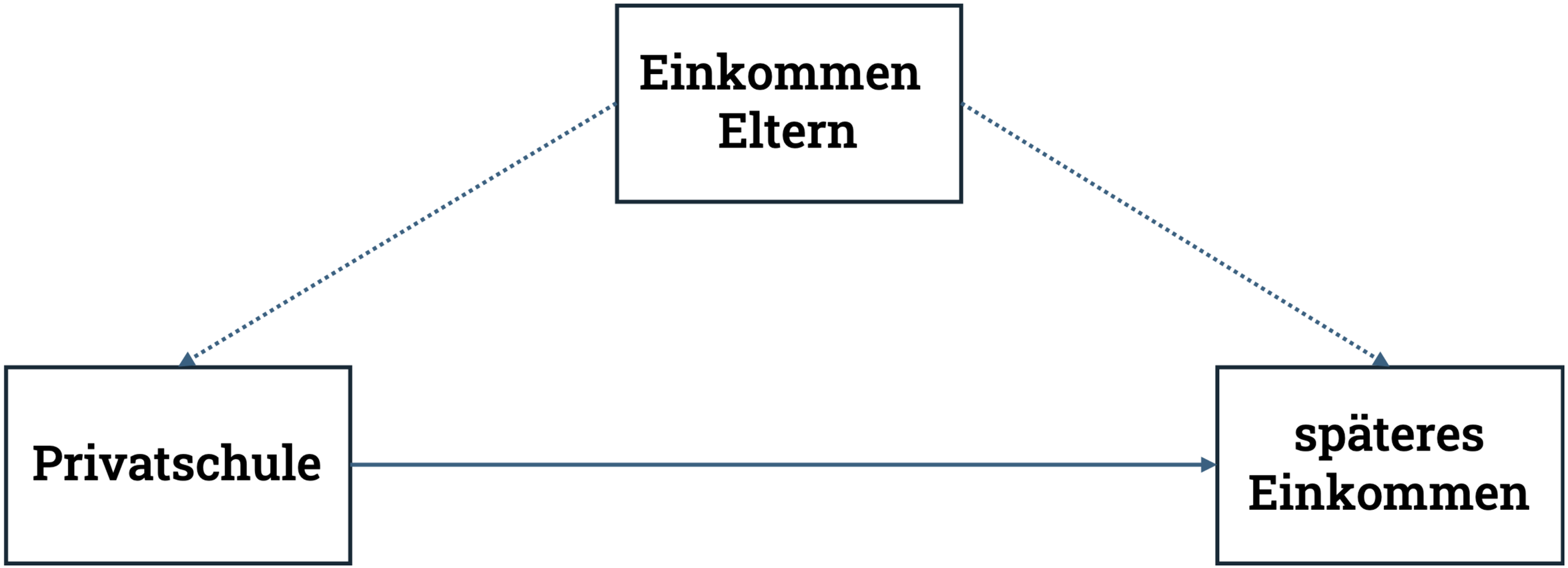


KAUSALANNAHMEN, CONFOUNDERS, COLLIDERS



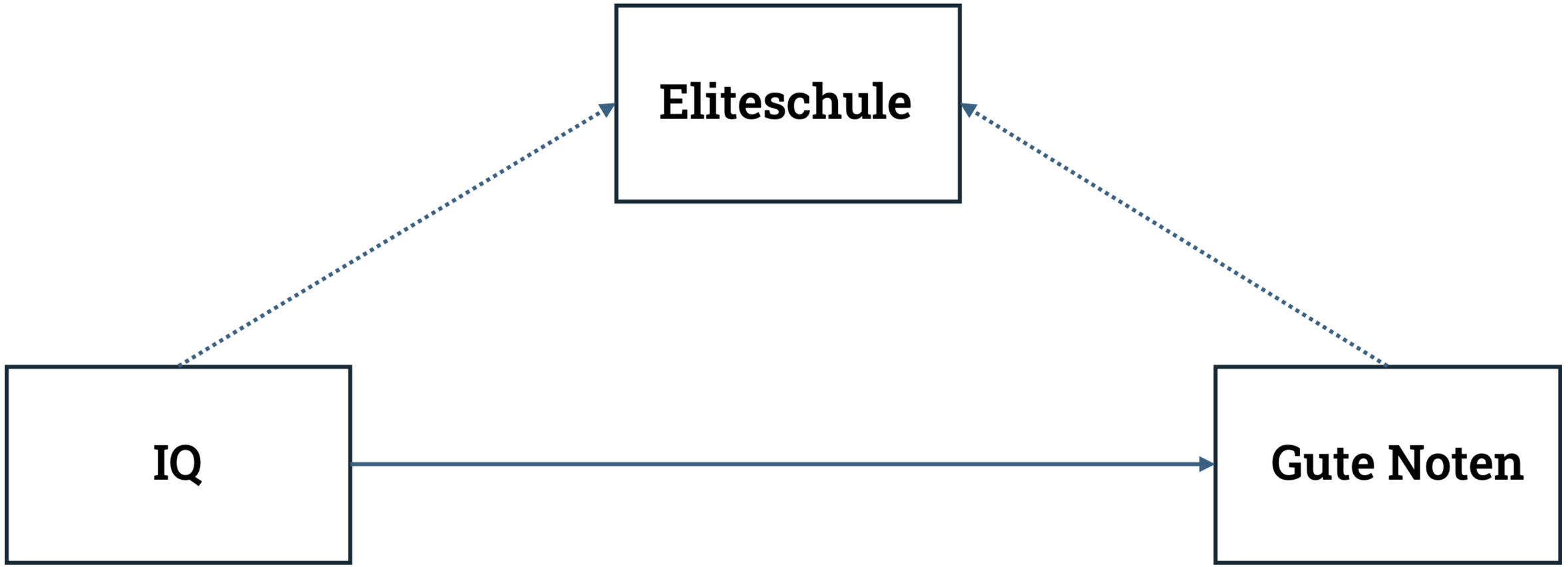
Quelle: https://catalogofbias.org

CONFOUNDER- ODER OMMITTED-VARIABLE-BIAS



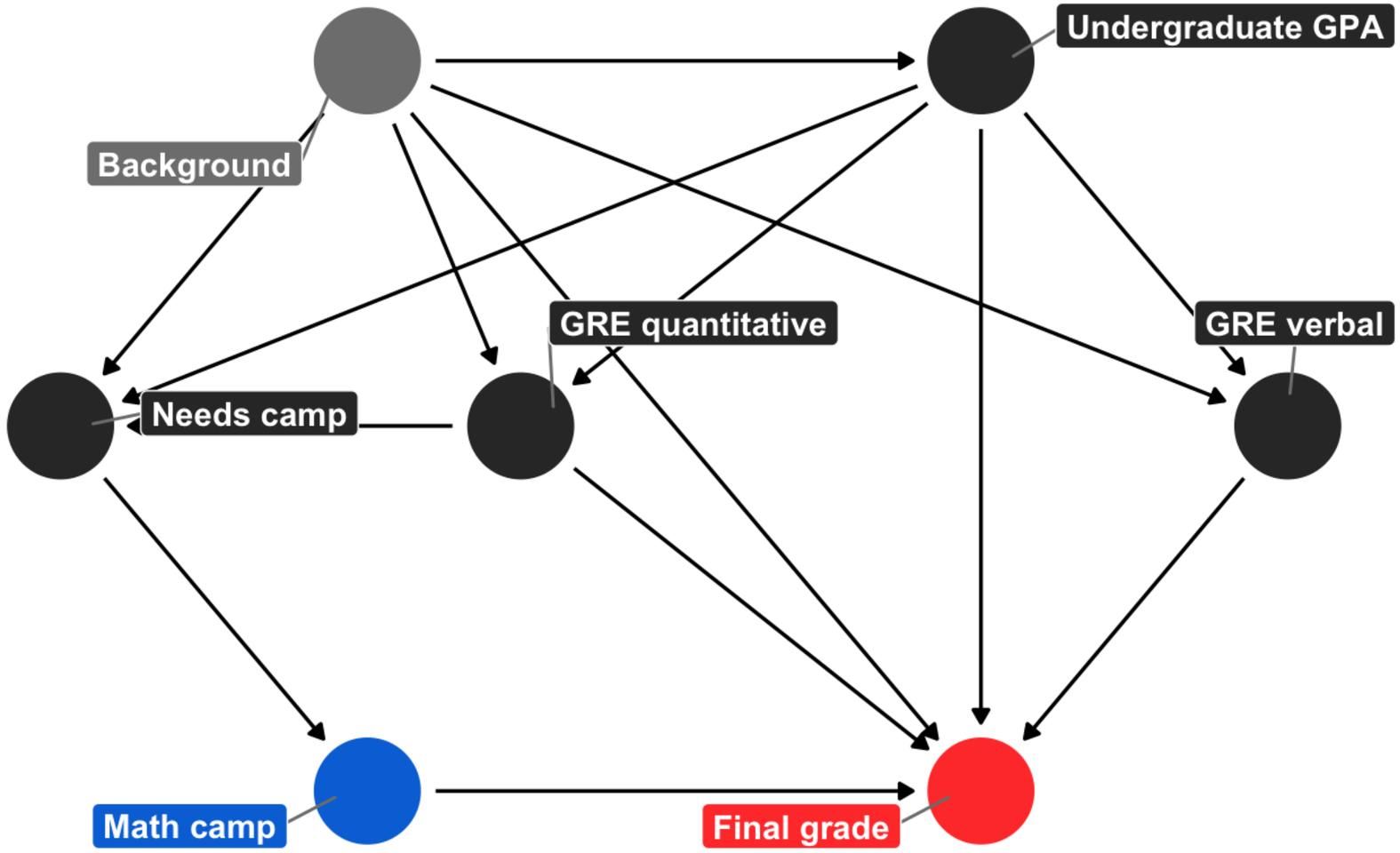
Nicht-Berücksichtigung einer relevanten Kovariate, die X *und* Y beeinflusst, verzerrt den geschätzen Zusammenhang zwischen X und Y.

COLLIDER-BIAS



Berücksichtigung einer Kovariate, die von X *und* Y beeinflusst wird, verzerrt den geschätzen Zusammenhang zwischen X und Y.

(KAUSALE) PFADMODELLE



Quelle: https://www.andrewheiss.com/blog/2020/02/25/closing-backdoors-dags/

VERLETZUNG DER MODELLANNAHMEN - UND NUN?

- Keine Panik! Modellannahmen sind praktisch immer verletzt (z.B. Normalverteilung der Residuen)
- viele Annahmen beziehen sich auf die Residuen, nicht auf X oder Y
- wichtig ist, einschätzen zu können, welche Konsequenzen eine Verletzung der Modellannahme haben kann
  - verzerrte Schätzer (zu hoch, zu niedrig)
  - falsche Standardfehler (Alpha- und Beta-Fehler)
  - falsche Kausalschlüsse (Rohrer, 2018; Coenen, 2022)
- vorsichtig formulieren, Robustheit der Ergebnisse prüfen

LITERATUR

Coenen, L. (2022). The indirect effect is omitted variable bias. A cautionary note on the theoretical interpretation of products-of-coefficients in mediation analyses. European Journal of Communication, 37(6), 679-688.

van Erkel, P. F., & Van Aelst, P. (2021). Why don’t we learn from social media? Studying effects of and mechanisms behind social media news use on general surveillance political knowledge. Political Communication, 38(4), 407-425.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. Advances in methods and practices in psychological science, 1(1), 27-42.