

Angewandte Statistik

Aaron Lohmann

Kausale Inferenz

- Neues Modul: “Modern Economics” im individuellen Ergänzungsbereich ab Wintersemester
- Gemeinsam mit Fiona Borsetzky
- 5 ECTS
- Besprechen und gemeinsames Lesen von wirtschaftswissenschaftlichen Forschungspapieren¹
- Soll Helfen dabei zu verstehen, was heutzutage in der VWL Forschung passiert.
- Super Vorbereitung für die Bachelorarbeit.

¹Das klingt viel, viel schlimmer als es ist! Im Gegenteil: Das kann richtig spannend sein!

- Modelle für unterschiedlich verteilte abhängige Variablen
 - Lineares Modell: Normalverteilung
 - Poisson: Zählvariablen
 - Logistisch: Binäre Variablen

- Modelle für unterschiedlich verteilte abhängige Variablen
 - Lineares Modell: Normalverteilung
 - Poisson: Zählvariablen
 - Logistisch: Binäre Variablen

Lernziel: Es ist wichtig über die abhängige Variable sorgfältig nachzudenken.

- Nichparametrische Methoden
- Einführung zu Machine Learning

- Nichtparametrische Methoden
- Einführung zu Machine Learning

Lernziel: Bias-Varianz Tradeoff, Grundideen von modernen Methoden.

Etwas vereinfacht ging es bisher immer um die Frage wie können wir Y angemessen durch X erklären. (Korrelation)

In den nächsten zwei Vorlesungen

- Kausale Inferenz als Modell, um über die Welt nachzudenken (heute)
- Kausale Inferenz in der Praxis mit `differnces-in-differences`
- Danach: Q&A und Zusammenfassung der Vorlesung.

Kausale Inferenz in zwei Begriffen:

- Inferenz: Das schließen auf Zusammenhänge unter Berücksichtigung von Information (Daten).

Kausale Inferenz in zwei Begriffen:

- Inferenz: Das schließen auf Zusammenhänge unter Berücksichtigung von Information (Daten).
- Kausalität: Wenn A dann B.

Kausale Inferenz in zwei Begriffen:

- Inferenz: Das schließen auf Zusammenhänge unter Berücksichtigung von Information (Daten).
- Kausalität: Wenn A dann B.

Conditio sine qua non (lat. „Bedingung ohne die nicht“):

Jede Handlung ist kausal, die nicht hinweggedacht werden kann, ohne dass der Erfolg in seiner konkreten Gestalt entfiel.

Zwei sinnvolle Definitionen:

1. Kausalität als Zusammenhang zwischen Ursache und Wirkung. (Wenn A dann B)

Zwei sinnvolle Definitionen:

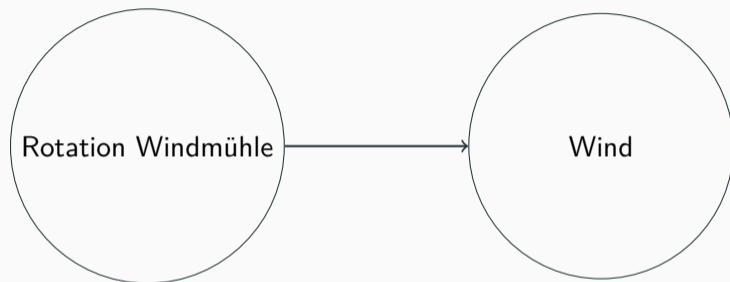
1. Kausalität als Zusammenhang zwischen Ursache und Wirkung. (Wenn A dann B)
2. Kausalität als Gedankenexperiment: Was-wäre-wenn Ansatz.

Zwei sinnvolle Definitionen:

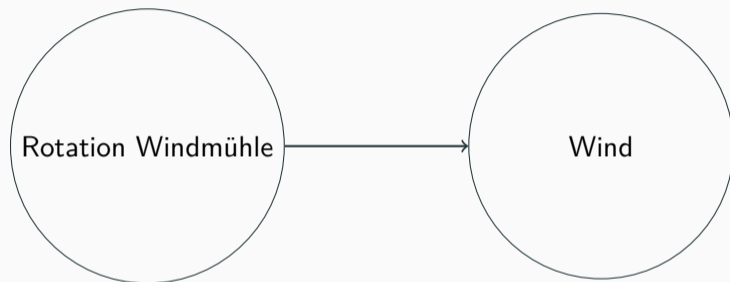
1. Kausalität als Zusammenhang zwischen Ursache und Wirkung. (Wenn A dann B)
2. Kausalität als Gedankenexperiment: Was-wäre-wenn Ansatz.

Wir werden sehen, dass die zweite Definition besonders hilfreich ist.

Korrelation vs. Kausalität - Beispiel 1: Windmühlen

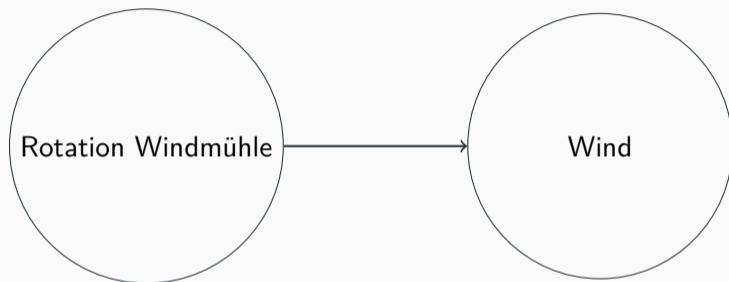


Korrelation vs. Kausalität - Beispiel 1: Windmühlen



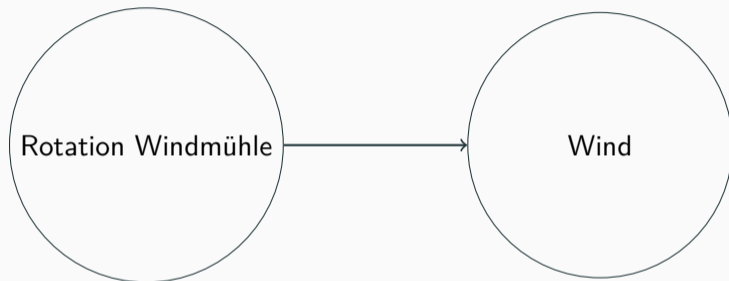
- Korreliert?

Korrelation vs. Kausalität - Beispiel 1: Windmühlen



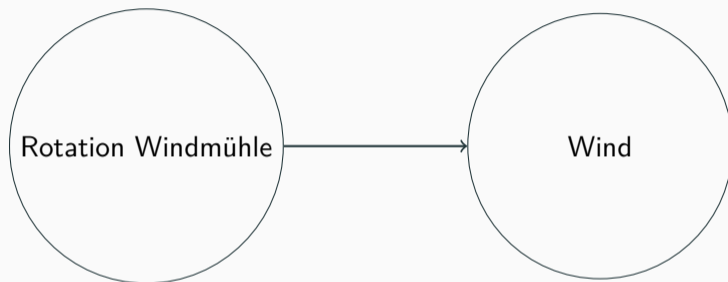
- Korreliert? Ja; typischerweise beobachten wir starke Rotation und Wind gemeinsam.

Korrelation vs. Kausalität - Beispiel 1: Windmühlen



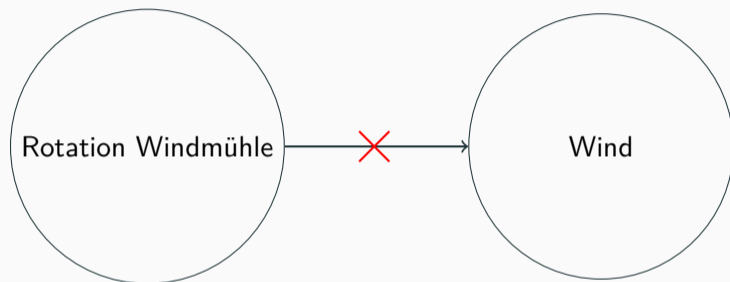
- Korreliert? Ja; typischerweise beobachten wir starke Rotation und Wind gemeinsam.
- Kausal?

Korrelation vs. Kausalität - Beispiel 1: Windmühlen



- Korreliert? Ja; typischerweise beobachten wir starke Rotation und Wind gemeinsam.
- Kausal?

Korrelation vs. Kausalität - Beispiel 1: Windmühlen



- Korreliert? Ja; typischerweise beobachten wir starke Rotation und Wind gemeinsam.
- Kausal? Nein; Die Rotation der Windmühle verursacht nicht den Wind.

Korrelation vs Kausalität - Beispiel 1: Windmühlen

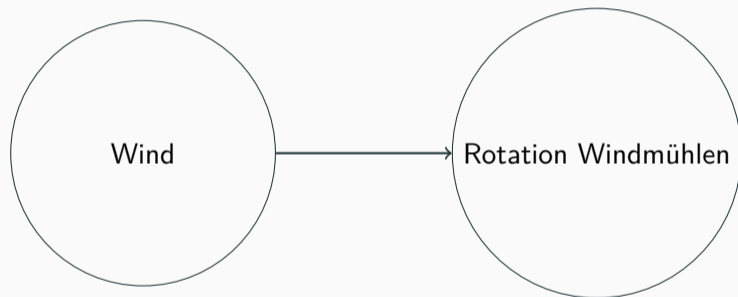
Woher wissen wir, dass die Rotation der Windmühle nicht den Wind verursacht?

²Bitte fragen Sie mich nicht nach Details – ich habe keine Ahnung

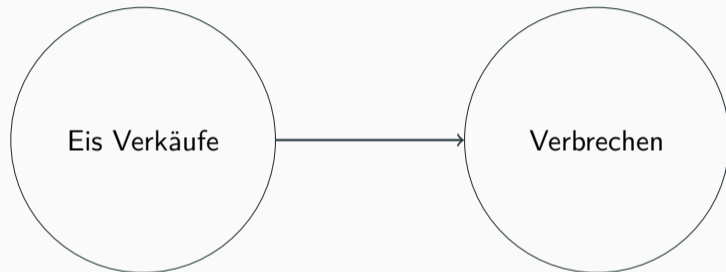
Woher wissen wir, dass die Rotation der Windmühle nicht den Wind verursacht?

*Physik!*²

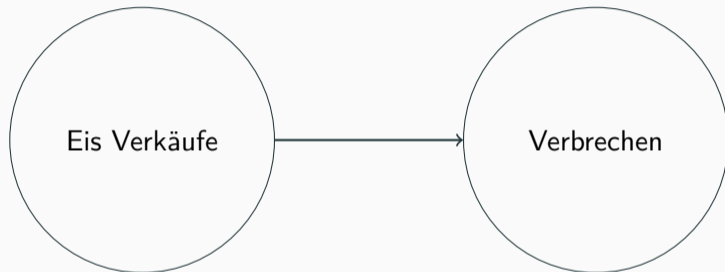
²Bitte fragen Sie mich nicht nach Details – ich habe keine Ahnung



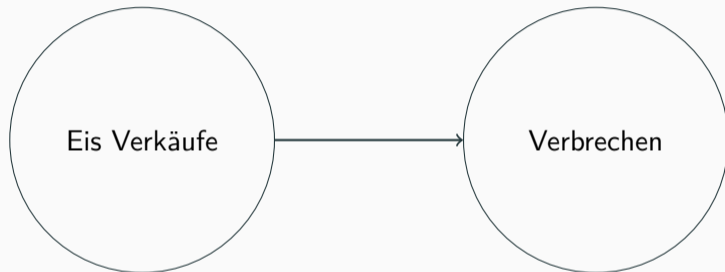
- Wir beobachten eine Korrelation zwischen Rotation und Wind und zwischen Wind und Rotation. Korrelation ist per Definition ungerichtet.
- Um eine kausale Beziehung herzustellen nutzen wir Wissen aus der Physik – ein Modell.



- Korreliert?



- Korreliert? Ja; wirkt aber komisch. Sollen wir nun Eis verkaufen verbieten?
- Kausal?



- Korreliert? Ja; wirkt aber komisch. Sollen wir nun Eis verkaufen verbieten?
- Kausal? Nein; Wahrscheinlich nicht, wir müssen nochmal nachdenken.

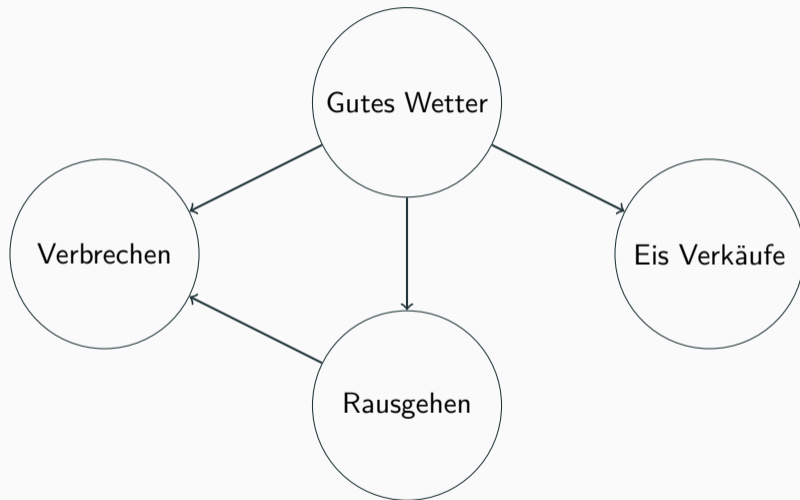
- Wann verkaufen wir Eis?

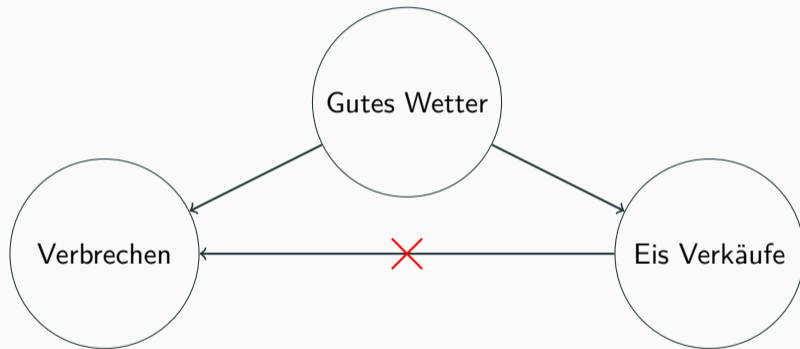
- Wann verkaufen wir Eis? Bei gutem Wetter.
- Wann gehen Menschen raus?

- Wann verkaufen wir Eis? Bei gutem Wetter.
- Wann gehen Menschen raus? Bei gutem Wetter.
- Wann passieren Verbrechen?

- Wann verkaufen wir Eis? Bei gutem Wetter.
- Wann gehen Menschen raus? Bei gutem Wetter.
- Wann passieren Verbrechen? Mit höherer Wahrscheinlichkeit, wenn sich Menschen draußen aufhalten.

Korrektur des Kausalen Graphen

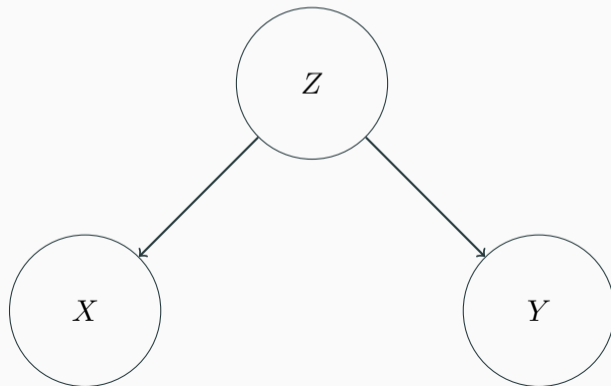




- Korrelation ist nicht gleich Kausalität.
- Hier ist das Wetter ein sogenannter confounder

Definition Confounder

Ein Confounder ist eine Variable, die sowohl mit der unabhängigen als auch der abhängigen Variable korreliert ist. Confounder können die Ergebnisse stark verzerren und sollten daher in der Analyse berücksichtigt werden.



Kausalität ohne Korrelation

- Nicht besonders intuitiv, aber durchaus relevant.
- Beispiel:
 - Schiffsruder und Richtung des Boots

Kausalität ohne Korrelation

- Nicht besonders intuitiv, aber durchaus relevant.
- Beispiel:
 - Schiffsruder und Richtung des Boots
 - Wenn das Schiff in einer geraden Linie fährt und die Kapitänin gegen den Wind steuert, dann sind Ruder und Richtung des Boots nicht korreliert, aber kausal.

- Nicht besonders intuitiv, aber durchaus relevant.
- Beispiel:
 - Schiffsruder und Richtung des Boots
 - Wenn das Schiff in einer geraden Linie fährt und die Kapitänin gegen den Wind steuert, dann sind Ruder und Richtung des Boots nicht korreliert, aber kausal.
 - Zentralbank und Inflation

- Nicht besonders intuitiv, aber durchaus relevant.
- Beispiel:
 - Schiffsruder und Richtung des Boots
 - Wenn das Schiff in einer geraden Linie fährt und die Kapitänin gegen den Wind steuert, dann sind Ruder und Richtung des Boots nicht korreliert, aber kausal.
 - Zentralbank und Inflation
 - Wenn die Zentralbank die Inflation stabil hält, dann sind Leitzinsen und Inflation nicht korreliert, aber kausal.

Takeaway des ersten Teils

- Korrelation und Kausalität sind nicht das Gleiche, kann gemeinsam auftreten, aber muss nicht.
- Für Kausalität brauchen wir ein Modell. Dieses Modell kann abstrakt sein von den Daten und mag auf vorherigen Wissen basieren.

Diskussion auf der nächsten WG-Party

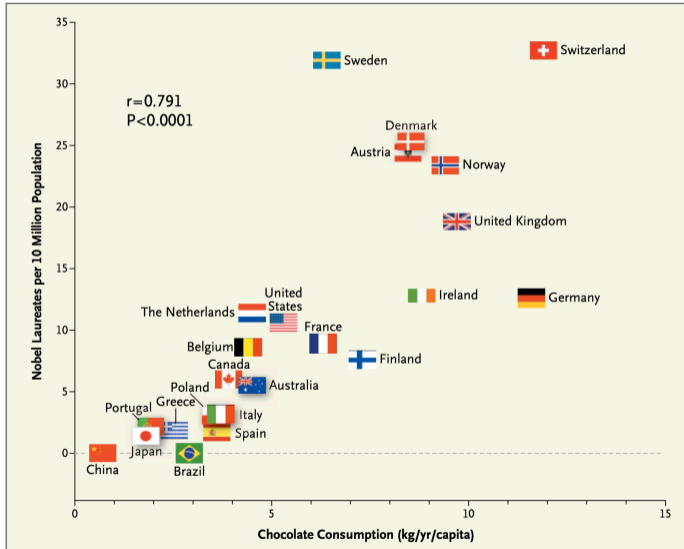


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

- Ein veröffentlichtes Paper...

Diskussion auf der nächsten WG-Party

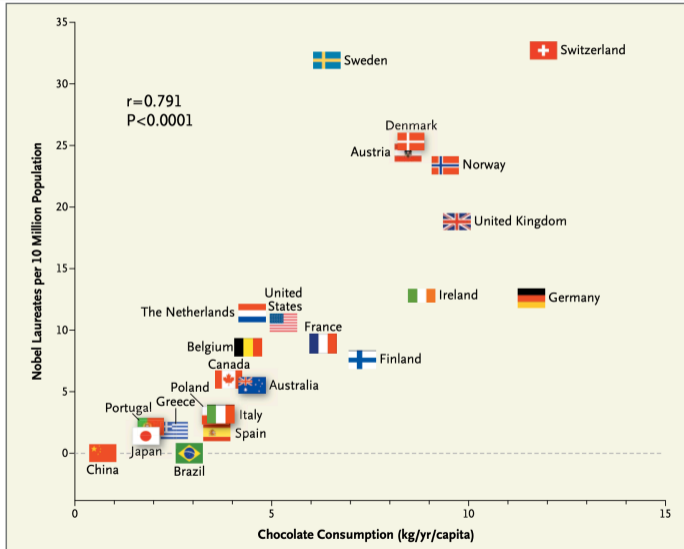


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

- Ein veröffentlichtes Paper...
- Fällt Ihnen ein kausaler Graph ein?

- Eine Art und Weise über Kausalität nachzudenken und etwas zu formalisieren.
- Einführung anhand eines Beispiels.

Potential Outcomes Framework - Beispiel (von Brady Neal)



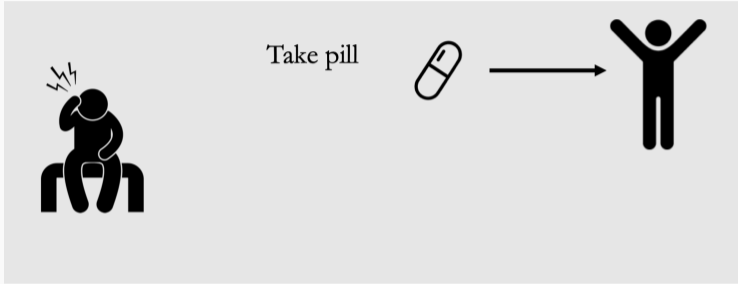
- Wir beobachten eine Person mit Kopfschmerzen.

Potential Outcomes Framework - Beispiel (von Brady Neal)



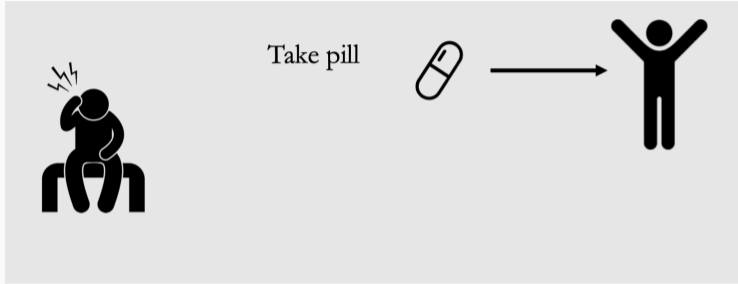
- Wir beobachten eine Person mit Kopfschmerzen.
- Wollen herausfinden, ob eine Pille hilft.

Potential Outcomes Framework - Beispiel Fortsetzung (von Brady Neal)



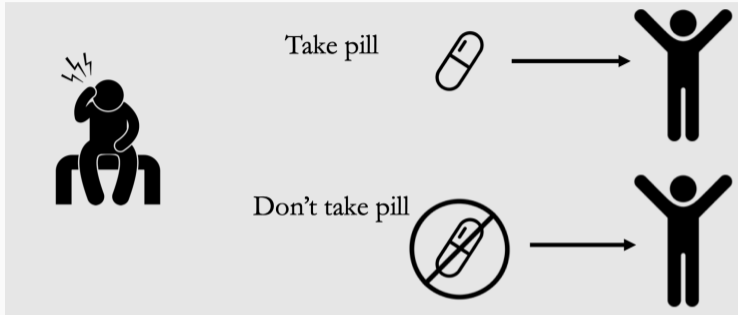
- Die Person nimmt eine Pille und ihr geht es wieder gut. Kausal?

Potential Outcomes Framework - Beispiel Fortsetzung (von Brady Neal)



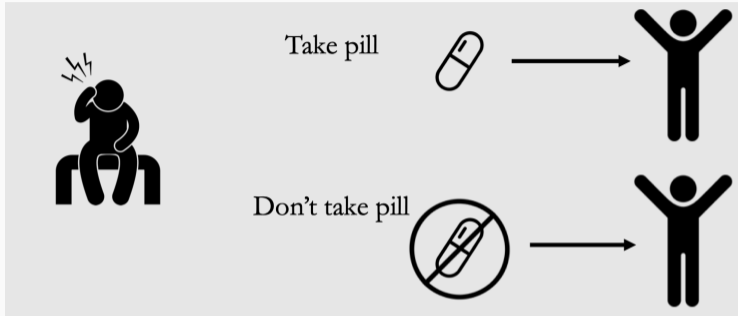
- Die Person nimmt eine Pille und ihr geht es wieder gut. Kausal?
- Nein; so einfach ist es leider nicht. Wir wissen nicht was passiert wäre, wenn die Person keine Pille genommen hätte.

Potential Outcomes Framework - Beispiel Fortsetzung (von Brady Neal)



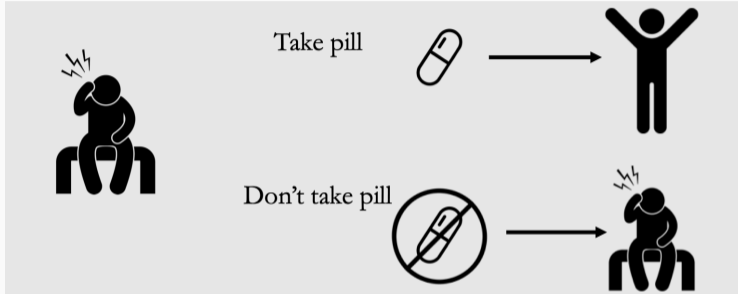
- Jetzt bewegen wir uns in unser kausales Gedankenexperiment und beobachten die Person mit und ohne Pille. Kausal?

Potential Outcomes Framework - Beispiel Fortsetzung (von Brady Neal)



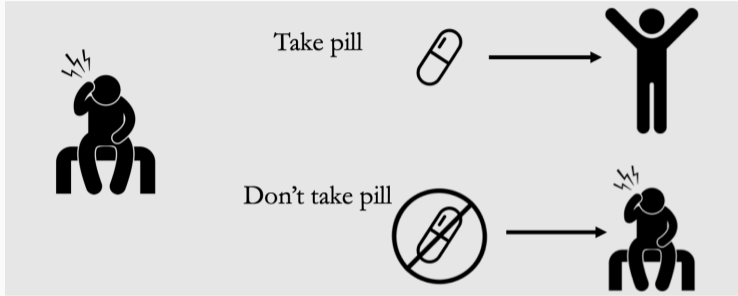
- Jetzt bewegen wir uns in unser kausales Gedankenexperiment und beobachten die Person mit und ohne Pille. Kausal?
- Wir können den Effekt schätzen, aber hier scheint die Pille nicht zu helfen, denn der Person geht es auch so wieder gut.

Potential Outcomes Framework - Beispiel Fortsetzung (von Brady Neal)



- Jetzt bewegen wir uns in unser kausales Gedankenexperiment und beobachten die Person mit und ohne Pille. Kausal?

Potential Outcomes Framework - Beispiel Fortsetzung (von Brady Neal)



- Jetzt bewegen wir uns in unser kausales Gedankenexperiment und beobachten die Person mit und ohne Pille. Kausal?
- Ja, die Pille hilft, denn die Person ohne Pille hat weiterhin Kopfschmerzen.

- Nun formalisieren wir das Ganze unter der Verwendung des Potential Outcomes Frameworks.
- Abhängige Variable: Y (z.B. Verbrechen im vorigen Beispiel)
- Behandlung (Treatment): D (z.B. Eisverkauf)

- $Y_i(1)$: Der Wert der abhängigen Variable Y für Person i , wenn sie die Behandlung erhält.
- $Y_i(0)$: Der Wert der abhängigen Variable Y für Person i , wenn sie die Behandlung nicht erhält.

Wie können wir mit $Y_i(0)$ und $Y_i(1)$ den kausalen Effekt von D auf Y bestimmen?

Wie können wir mit $Y_i(0)$ und $Y_i(1)$ den kausalen Effekt von D auf Y bestimmen?

- $Y_i(1) - Y_i(0)$: Der kausale Effekt von D auf Y für Person i .

Wie können wir mit $Y_i(0)$ und $Y_i(1)$ den kausalen Effekt von D auf Y bestimmen?

- $Y_i(1) - Y_i(0)$: Der kausale Effekt von D auf Y für Person i .
- Was geht hier schief?

Wie können wir mit $Y_i(0)$ und $Y_i(1)$ den kausalen Effekt von D auf Y bestimmen?

- $Y_i(1) - Y_i(0)$: Der kausale Effekt von D auf Y für Person i .
- Was geht hier schief?
- Wir können $Y_i(1)$ und $Y_i(0)$ nicht gleichzeitig beobachten.

Das fundamentale Problem der Kausalität

Direktes beobachten von kausalen Effekten ist unmöglich, da wir nur eine der beiden Potential Outcomes beobachten können.

- Wir können nicht eine Person mehrmals beobachten, aber wir können viele Personen beobachten.

- Wir können nicht eine Person mehrmals beobachten, aber wir können viele Personen beobachten.
- Dann können wir doch einfach den Durchschnitt bilden, oder?

- Wir können nicht eine Person mehrmals beobachten, aber wir können viele Personen beobachten.
- Dann können wir doch einfach den Durchschnitt bilden, oder? Genau!
- Das ist die Idee hinter dem `average treatment effect` (ATE). Oder auf Deutsch: Der durchschnittliche Behandlungseffekt.

Average Treatment Effect (ATE)

- Formal also: $ATE = E[Y_i(1) - Y_i(0)]$

Average Treatment Effect (ATE)

- Formal also: $ATE = E[Y_i(1) - Y_i(0)]$
- Unter bestimmten Annahmen können wir den ATE schätzen.

Average Treatment Effect (ATE)

- Formal also: $ATE = E[Y_i(1) - Y_i(0)]$
- Unter bestimmten Annahmen können wir den ATE schätzen.

$$ATT = E[Y_i(1)] - E[Y_i(0)] = E[Y_i(1)|D = 1] - E[Y_{-i}(0)|D = 0]$$

- Oder etwas konkreter: Wir haben zwei Gruppen, eine die die Behandlung erhält und eine die die Behandlung nicht erhält.

$$ATT = \sum_{j=1}^{N_1} \frac{Y_j(1)}{N_1} - \sum_{i=1}^{N_0} \frac{Y_i(0)}{N_0}$$

- N_1 und N_0 sind die Anzahl der Personen in den jeweiligen Gruppen.

Erklärung zur vorherigen slides

- Wir nutzen bilden Mittelwerte für die Gruppen, die die Behandlung erhalten haben und die die die Behandlung nicht erhalten haben.
- Dann subtrahieren wir die beiden Mittelwerte.
- Eine wichtige Annahme ist, dass die Gruppen gleich wahrscheinlich der Behandlung ausgesetzt sind. Formal also

$$(Y_1, Y_0) \perp D$$

das Symbol \perp steht für unabhängig. Die Behandlung ist unabhängig von den Potential Outcomes. In anderen Worten: Die Bhandlung ist zufällig.

Erklärung zur vorherigen slides

- Wir nutzen bilden Mittelwerte für die Gruppen, die die Behandlung erhalten haben und die die die Behandlung nicht erhalten haben.
- Dann subtrahieren wir die beiden Mittelwerte.
- Eine wichtige Annahme ist, dass die Gruppen gleich wahrscheinlich der Behandlung ausgesetzt sind. Formal also

$$(Y_1, Y_0) \perp D$$

das Symbol \perp steht für unabhängig. Die Behandlung ist unabhängig von den Potential Outcomes. In anderen Worten: Die Bhandlung ist zufällig.

Wir schauen uns das nicht formal an, sondern in einem Beispiel.

siehe code

Unabhängigkeitsannahme mit Kovariaten

- In der Praxis ist die Unabhängigkeitsannahme oft nicht erfüllt.
- Wir können die Unabhängigkeitsannahme mit Kovariaten verbessern.

$$(Y_1, Y_0) \perp D | X$$

- Kausalität ist nicht gleich Korrelation, manchmal schon und manchmal gibt es keine Korrelation trotz Kausalität.
- Wenn wir Kausalität schätzen wollen, dann hilft es im Rahmen eines Modells strukturiert über die Welt nachzudenken.
- Das Potential Outcomes Framework ist ein solches Modell, das uns hilft Kausalität zu schätzen. Dabei haben wir gelernt, dass man idealerweise mehrere Beobachtungen von der gleichen Person hätte.

- Kausale Inferenz Modelle versuchen häufig so nah wie möglich an eben das Potential Outcomes Framework zu kommen.
- Man versucht Gruppen zu bilden, die gut vergleichbar sind.
- Nächste Woche eine dieser Methoden: `differences-in-differences`

Fragen?