Kausale Inferenz

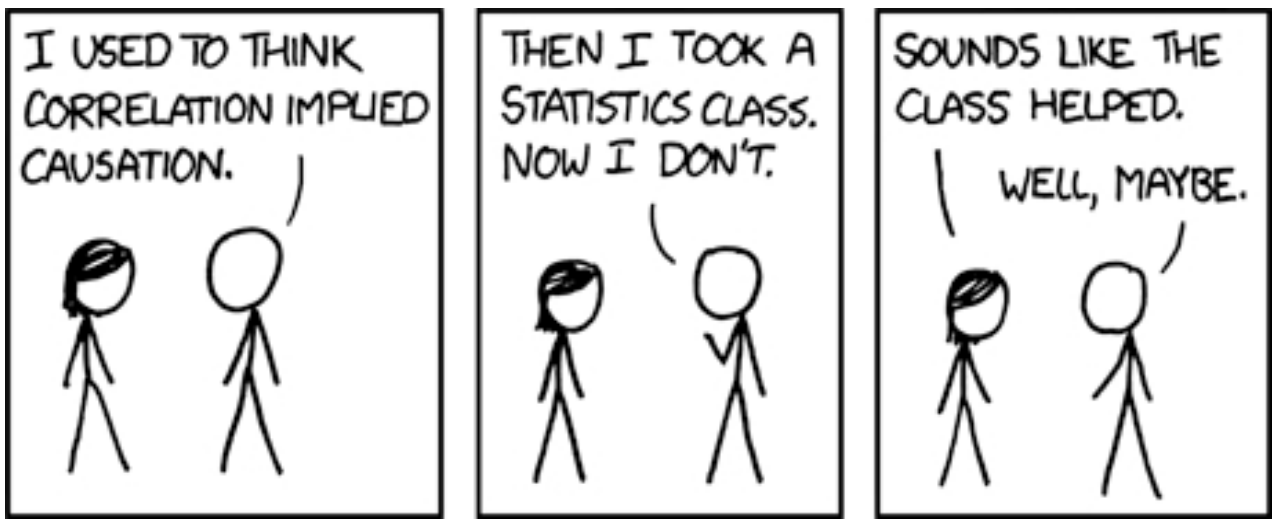
Kapitel 1: Einführung

Potential Outcomes & Kausale Effekte

Dezember 2019
Version: 10. Dezember 2019

Michael Schomaker
UMIT University, Institute of Public Health, MDM and HTA

Was wisst Ihr über kausale Inferenz?



Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Inhalt – Tag 1

- Kapitel 1:
  - Definition von kausalen Effekten, potential outcomes, etc.
  - die kausale “road map”
- Kapitel 2:
  - das kausale Modell $\rightarrow$ Hintergrundwissen in DAGs
  - Grundkonzepte: Collider, Confounder, Mediator, back-door Pfad etc.
- Kapitel 3: Die wichtigsten Annahmen, um mit Kausalität zu arbeiten

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Inhalt – Tag 2

- Kapitel 3: “Identifizierbarkeit”
- Kapitel 4: randomisierte Studien und Experimente (und deren Probleme)
- Kapitel 5: Schätzung kausaler Effekte mit Standardisierung

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Inhalt – Tag 3

- Kapitel 6: Schätzung kausaler Effekte mit “inverse probability weighting”
- Kapitel 7: Regression, Kollapsibilität & Effektmodifikation
- Kapitel 8: Ausblick

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Vorkenntnisse

- lineares und logistisches Regressionsmodell
- ...?

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Literatur

- M. Hernan and J. Robins. *Causal inference*. Chapman & Hall/CRC, Boca Raton, 2020.
  https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
- S. L. Morgan and C. Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, 2015.
- J. Pearl. *The Book of Why*. Basic Books, 2018.
- G. Imbens and and D. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

...und die vielen Referenzen zu Kapitelende

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Was ist eine kausale Fragestellung?

Intuitiv klar!

BMJ 2016;355:i6536 doi: 10.1136/bmj.i6536 (Published 9 December 2016)

FEATURE

CHRISTMAS 2016: FOOD FOR THOUGHT

Is caviar a risk factor for being a millionaire?

Anders Huitfeldt argues that the answer depends on your definition of "risk factor" and calls for greater clarity in research

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

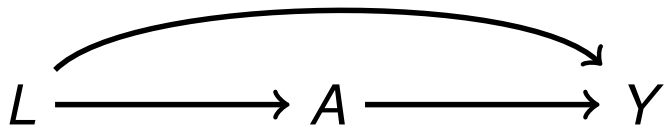
Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

- *"Most studies in the health, social and behavioral sciences aim to answer **causal** rather than associative – questions. Such questions require some knowledge of the data-generating process, and cannot be computed from the data alone, nor from the distributions that govern the data."* (Judea Pearl)

- *"In fact, had you not understood [these] causal concepts, you would have not survived long enough to read this chapter – or even to learn to read"* (Miguel A. Hernán)

Assoziation versus Kausalität

Birthdays are good for you! Statistics show that people who have the most live the longest.
Larry Lorenzoni



“Assoziation heißt nicht Kausalität”
“Korrelation beinhaltet nicht Kausalität”

$\rightarrow$ Die Richtung ist entscheidend!

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Klassische Statistik

- Assoziation: Jeglicher Zusammenhang, der über die gemeinsame Verteilung der beobachteten Daten definiert wird (z.B. Korrelationskoeffizient, Regression, χ^2-Unabhängigkeitstest)

- Regressionsmodelle:

  - deskriptiv [=assoziativ?]

  - prädiktiv

  - ...kausal interpretierbar?

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Kausale Statistik

Kausale Frage
[kontrafaktisch formuliert, Zielparameter ψ, $\rightarrow$ Kapitel 1]

$\downarrow$

Kausales Modell
[Hintergrundwissen, $\rightarrow$ Kapitel 2]

$\downarrow$

Identifizierung:
[ψ schätzbar, auf Basis von kausalem Modell und weiteren Annahmen? $\rightarrow$ Kapitel 3]

$\downarrow$

Statistisches Modell:
[Wenn ψ identifizierbar, dann wähle geeignetes statische Methodik aus $\rightarrow$ Kapitel 4 ff.]

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Detailliertere “road map” [1]

1. Specify knowledge about the system to be studied using a causal model.

Represent background knowledge about the system to be studied. A causal model describes the set of possible data-generating processes for this system.

2. Specify the observed data and their link to the causal model.

Specify what variables have been or will be measured, and how these variables are generated by the system described by the causal model.

3. Specify a target causal quantity.

Translate the scientific question into a formal causal quantity (defined as some parameter of the distribution of counterfactual random variables), following the process in Figure 3.

4. Assess identifiability.

Assess whether it is possible to represent the target causal quantity as a parameter of the observed data distribution (estimand), and, if not, what further assumptions would allow one to do so.

5. State the statistical estimation problem.

Specify the estimand and statistical model. If knowledge is sufficient to identify the causal effect of interest, commit to the corresponding estimand. If not, but one still wishes to proceed, choose an estimand that under minimal additional assumptions would equal or approximate the causal effect of interest.

6. Estimate.

Estimate the target parameter of the observed data distribution, respecting the statistical model.

7. Interpret.

Select among a hierarchy of interpretations, ranging from purely statistical to approximating a hypothetical randomized trial.

FIGURE 1. A general roadmap for approaching causal questions.

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

[1] aus Petersen und van der Laan [1]

Potential Outcomes

"The fundamental notion underlying our approach is that causality is tied to an *action* (or manipulation, treatment or intervention), applied to a *unit*. [...] a causal statement presumes that, although a unit was [...] subject to, or exposed to, a particular action, treatment or regime, the same unit could have been exposed to an alternative action treatment or regime. [...] For instance, when deciding to take an aspirin to relieve your headache, you could have also chosen not to take the aspirin, or an alternative medicine. [...] Given a unit and a set of actions, we associate each action-unit pair with a *potential outcome*. We refer to these outcomes as potential outcomes because only one will ultimately be [...] observed. Ex post, the other potential outcomes cannot be observed because the corresponding actions that would lead to them being realized were not taken."[2]

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

[2] from Imbens and Rubin [2]

Terminologie

- Informell:
  - Ursache (A) $\rightarrow$ Wirkung (Y)
  - Wie stark ist der *Effekt* von A auf Y?
- Etwas genauer: Wenn alle Personen statt $A = a'$ nun $A = a''$ bekommen würden, wie würde sich dann $E(Y)$ verändern?
- Potential Outcome: $Y^{a'}$ = Zufallsvariable Y, die beobachtet worden wäre, wenn alle Personen der Population $A = a'$ bekommen hätten.

  $\rightarrow$ "kontrafaktisch", da nicht jeder zwangsläufig $A = a'$ als Realisierung hat

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Terminologie

Y = Zielgröße
A = Intervention (Exposure, Treatment)
L = Andere zur Verfügung stehende gemessene Variablen
U = ungemessene Variablen

Y^a = kontrafaktische Zielgröße für $A = a$

Wenn A binär ist besteht folgender Zusammenhang:

$$\begin{aligned} Y &= Y^1 \quad \text{wenn} \quad A = 1 \\ Y &= Y^0 \quad \text{wenn} \quad A = 0 \end{aligned}$$

und damit

$$Y = AY^1 + (1 - A)Y^0 \qquad (1)$$

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Potential Outcomes = fehlende Daten

i	A	Y	Y^0	Y^1
1	0	2.3	2.3	?
2	0	4.8	4.8	?
3	0	0.5	0.5	?
4	0	1.2	1.2	?
5	1	5.1	?	5.1
6	1	3.9	?	3.9
7	1	4.6	?	4.6
8	1	2.0	?	2.0

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Kausale Effekte sind oft

- zu vage definiert (“the effect of treatment”)
- oder so definiert, dass es eine Auswertung bequem macht (nimm Regressionsschätzungen)

Am besten: starte mit klarer Fragestellung und wähle erst dann den passenden Analyseansatz

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Kausale Effekte

Definition: kausaler Effekt

Ein individueller kausaler Effekt liegt vor, wenn

$$y_i^1 \neq y_i^0 . \tag{2}$$

Ein *Average Treatment Effect (ATE)* liegt vor, wenn

$$E(Y^1) \neq E(Y^0) . \tag{3}$$

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Kausale Zielgrößen / “target quantity”

Sei Y binär; dann sind typischerweise die folgenden (kausalen) Größen von Interesse:

absolute Risikodifferenz: $\psi_{ATE} = P(Y^1 = 1) - P(Y^0 = 1)$

relatives Risiko: $\psi_{RR} = P(Y^1 = 1)/P(Y^0 = 1)$

Odds Ratio $\psi_{OR} = \frac{P(Y^1=1)/P(Y^1=0)}{P(Y^0=1)/P(Y^0=0)}$

Frage: was ist hier die Interpretation?

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Bedingte Effekte

Bedingt auf Kovariablen: z.B. $E(Y^1|L=l) - E(Y^0|L=l)$

$\rightarrow$ Hypothese der *Effektmodifikation*, also dass der kausale Effekt für z.B. $L=0$ und $L=1$ separat von Interesse ist .

Beispiel: Y soziale Entwicklung eines Kindes (Score), $A = 1/2$ aufgewachsen bei 1 bzw. 2 Elternteilen.

ATE: $E(Y^2|L_1, L_2) - E(Y^1|L_1, L_2)$. Effekt, stratifiziert für L_1=Geschlecht und L_2= Alter (ab dem nur noch bei einem Elternteil aufgewachsen), ist von Interesse.

Andere bedingte Effekte, z.B. *Average Treatment Effect of the Treated*: $E(Y^1|A=1) - E(Y^0|A=1)$

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Strukturelle kausale Modelle

- Eine andere Art um kausale Effekte zusammenzufassen
- Oft verwendet bei stetigen Interventionsvariablen
- Marginales strukturelles (kausales) Modell:

$$\begin{aligned} E(Y^a) &= \beta_0 + \beta_1 a \\ E(Y^a) &= \beta_0 + \beta_1 f(a) \end{aligned}$$

"Marginal", da nicht bedingt

"Strukturell" da kausale Abhängigkeitsannahmen der Variablen im Modell stecken ($\rightarrow$ Kapitel 2)

"kausal" da Y^a und nicht Y

- Auch bedingte Effekte potentiell von Interesse:

$$E(Y^a|L) = \beta_0 + \beta_1 a + \beta_2 L + \beta_3 aL$$

- "working models"

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Wie definiere ich die Intervention?

- Klare Definition unabdingbar um kausalen Effekt zu identifizieren ($\rightarrow$ Kapitel 3).

  Beispiel: $A = 1/2$ aufgewachsen bei 1 bzw. 2 Elternteilen. Was heißt das? Geschieden vs. nicht geschieden? Zweites Elternteil zu Besuch? Wieviel Jahre alleinerziehendes Elternteil?

- Was sind 1) die Interventionsvariablen und 2) wie wird auf die Interventionsvariablen interveniert?

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Auf welchen Variablen wird interveniert?

- 1 Variable an einem 1 Zeitpunkt.

  Beispiel: $A =$ Verwendung einer Pillenbox für antiretrovirale Medikamente, $Y =$ Unterdrückung der Virenlast [3]

- 1 Variable zu verschiedenen Zeitpunkten.

  Beispiel: $A =$ regelmäßige Medikamenteneinnahme (1,1,1,...,1) vs. keine Einnahme (0,0,0,...,0) [4]

- Mehrere Variablen (1 oder mehrere Zeitpunkte)

  Beispiel: $A =$ Kombination aus kein Rauchen, mehr Sport, verbesserte Ernährung und weniger Alkohol. $Y =$ Herzerkrankungen [5]

- Intervention auch auf fehlende Daten möglich: Zielgröße bei keinem Drop-out.

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Wie wird interveniert?

- statisch

  Beispiel: $A = 1$: nimm antiretrovirale Medikamente

- dynamisch [abhängig von L]

  Beispiel: $A = 1$: nimm antiretrovirale Medikamente wenn CD4 count (L_1) < 750 cells/mm^3

- stochastisch

  Beispiel: nimm antiretrovirale Medikamente mit Wahrscheinlichkeit 80%

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Interventionsvariablen[3]

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Decision 1: On which variables will you intervene?

Interventions on one exposure variable

Point treatment effects: Effect of exposure or intervention at a single time point.[47, 51]

Example: What is the effect of using a pill box during a given month on adherence to antiretroviral medications? [64]

Interventions on multiple exposure variables

Longitudinal treatment effects: Cumulative effect of multiple treatment decisions or exposure interventions over time.[2, 35, 36]

Example: Does cumulative exposure to iron supplementation over the course of pregnancy affect probability of anemia at delivery? [65]

Missing data, losses to follow up, and censoring: Effect of a point or longitudinal treatment when some data are missing or outcomes are not observed.[2, 48, 49, 66]

Example: What would the results of a randomized trial have looked like if study drop out had been prevented?

Direct and indirect effects: Portion of an effect that is or is not mediated by an intermediate variable.[19, 32-34]

Example: Would diaphragm and lubricant use reduce risk of HIV acquisition if their effect on condom use were blocked? [67]

[3] aus Petersen und van der Laan [1]

Art der Intervention[4]

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Decision 2: How will you set the values of these intervention variables?

Static intervention: Value of the exposure variable(s) set deterministically for all members of the population.

Example: How would mortality have differed if all HIV infected patients meeting WHO immunologic failure criteria had been switched to second line antiretroviral therapy immediately, versus if none of these patients had been switched? [68]

Dynamic regime/individualized treatment rule: Value of the exposure variable(s) assigned based on individual characteristics.[26-28]

Example: How would mortality have differed if subjects had been assigned to start antiretroviral therapy at different CD4 T cell count thresholds? [31]

Stochastic intervention: Value of the exposure variable(s) assigned from some random distribution.[29-31]

Example: How would a change in the distribution of exercise in a population affect risk of coronary heart disease? [29]

[4] aus Petersen und van der Laan [1]

Zielgröße / “target quantity” / ψ[5]

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Decision 3. What summary of counterfactual outcome distributions is of interest?

Absolute versus relative contrast: Causal risk difference versus relative risk versus odds ratio.

Effect defined using a marginal structural model: Model that describes how the expected counterfactual outcome varies as a function of the counterfactual static or dynamic intervention.[2, 27, 28, 35, 36]

Example: The counterfactual hazard of mortality as a function of time to switching treatment following virologic failure of antiretroviral therapy and elapsed time since failure is summarized using a logistic main term model. [69]

Marginal effect versus effect conditional on covariates: When the counterfactual outcome is a non-linear function of the exposure, or if the exposure interacts with covariates, these are different quantities.

Example: The marginal counterfactual hazard ratio defined using a marginal structural cox model is different from the counterfactual hazard ratio defined using a marginal structural cox model that also conditions on baseline covariates (due to the non-collapsibility of the hazard ratio).[70]

Effect modification: Comparison of the effect between strata of the population.

Example: How does effect of efavirenz versus nevirapine based antiretroviral therapy differ among men versus among women?[71]

[5] aus Petersen und van der Laan [1]

Population[6]

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Decision 4: What population is of interest?

Whole population: Causal effect in the population from which the observed data were drawn.

Example: What is the effect of diaphragm use on HIV acquisition in the target population from which trial participants were sampled?

Subset of the population: Causal effect in a subset of the target population from which the observed data were drawn. The subset might be defined based on covariates values (as when effect modification is evaluated), or on the population that actually received or did not receive the exposure (effect of treatment on the treated or untreated, respectively).[72]

Example: What was the effect of diaphragm use on HIV acquisition among those women who actually used a diaphragm?

A different population: Causal effect in some target population other than that from which the data were drawn ("transportability").[27, 43-45]

Example: What would the effect of diaphragm and gel use have been in a population in which the factors determining condom use were different?

[6] aus Petersen und van der Laan [1]

Übung 1

Definiere passende “target quantities” –formal korrekt– basierend auf folgenden (unpräzisen) Aussagen:

1. “Mich interessiert der Anteil an Todesfällen bei Patienten mit Speiseröhrenkrebs falls diese sowohl mit Chemo- als auch Radiotherapie behandelt werde.”
2. “Um wieviel *kg* nehmen Kinder im Mittel in einem Jahr zu, wenn sie Therapie A beginnen im Vergleich zu Therapie B?”
3. “Was ist der durchschnittliche BMI von übergewichtigen Patienten, nach einem Monat in dem mindestens 30 Minuten Bewegung pro Tag eingehalten wurden und auf Rauchen verzichtet wurde?”
4. “Wieviel wahrscheinlicher ist es mit Medikament *A* geheilt zu werden als mit Medikament *B*?”
5. “Wie wirkt sich die Dosis des Medikaments aus die Heilungsdauer (in Tagen) aus?”
6. “Unterscheidet sich der Effekt der Medikamentendosis auf die Heilungsdauer zwischen Rauchern und Nichtrauchern?”

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise

Literaturnachweise für Beispiele

[1] M. L. Petersen and M. J. van der Laan.
Causal models and learning from data: integrating causal modeling and statistical estimation.
Epidemiology, 25(3):418–26, 2014.

[2] G. Imbens and D. Rubin.
Causal Inference for Statistics, Social, and Biomedical Sciences.
Cambridge University Press, New York, USA, 2015.

[3] M. L. Petersen, Y. Wang, M. J. van der Laan, D. Guzman, E. Riley, and D. R. Bangsberg.
Pillbox organizers are associated with improved adherence to HIV antiretroviral therapy and viral suppression: A marginal structural model analysis.
Clinical Infectious Diseases, 45(7):908–915, 2007.

[4] D. Westreich, S. R. Cole, J. G. Young, F. Palella, P. C. Tien, L. Kingsley, S. J. Gange, and M. A. Hernan.
The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death.
Statistics in Medicine, 31(18):2000–2009, 2012.

[5] S. L. Taubman, J. M. Robins, M. A. Mittleman, and M. A. Hernan.
Intervening on risk factors for coronary heart disease: an application of the parametric g-formula.
International Jornal of Epidemiology, 38(6):1599–611, 2009.

Motivation
Definitionen
Potential Outcomes
Kausale Effekte
Interventionen
Zusammenfassung
Übung
Literaturnachweise