Causal Inference Theory & Practice

October 2023

Dimitra (Mimie) Liotsiou, dunnhumby

dunnhumby

“

dunnhumby is the global leader in Customer Data Science, empowering businesses everywhere to compete and thrive in the modern data-driven economy.
We always put the Customer First.

dunnhumby clients

North America

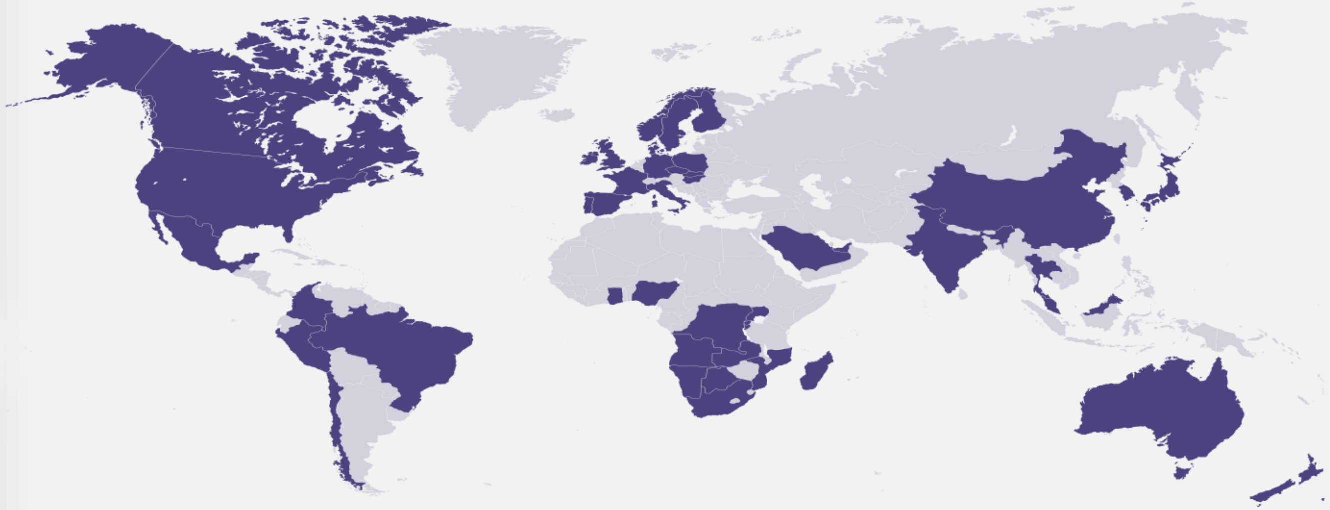
Raley's, McDonald's, weis, Southeastern Grocers, meijer, metro, Lowes Foods, SpartanNash, 7-Eleven

Latin America

COPSERVIR, SMU, éxito, RD Gente, Saúde e Bem-estar, GPA, Soriana, Farmacias Peruanas

Our CPG partners include:

Unilever, Nestlé, MARS, L'ORÉAL, Coca-Cola, Heinz, P&G, PEPSICO, General Mills, Kellogg's



UK & Ireland & EMEA

TESCO, coop, MONOPRIX, coop Norge, CRAI, BOOKER, coop, SHOPRITE, SOK

APAC & China

イトーヨーカドー, TESCO, foodstuffs, Lotus's, Vanguard 华润万家

78
Retail Clients

1250+
CPG Clients

770M
Active customers under management

30B
Baskets per year

Agenda

1. Theory basics

The causal data science pipeline: 4 stages

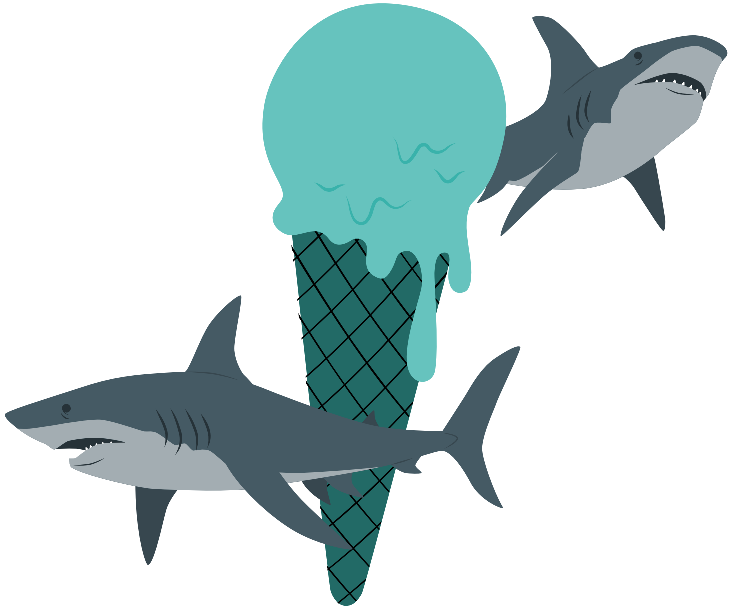
2. Practice

Quick tour of application to retail case study in Python

Correlation does not equal causation!

Spurious correlations

Number of shark attacks positively correlated with volume of ice cream sales at sampled beaches over the year



Cannot sensibly interpret these correlations as causation!

Country-wide chocolate consumption positively correlated with Nobel prizes per capita

Correlation vs. Causation

All DS/ Stats/ ML methods:
Correlations (or associations)

E.g. regression coefficients, $\widehat{\beta}_i$:

- Associations, incl. spurious. Bias.
- Not causal effect of X_i
- Effects of *interventions*

Causal questions require causal methods

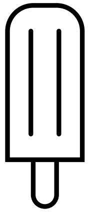
Some examples

What is the **effect,** or **impact,** of …

- This media campaign
- A customer joining a loyalty scheme
- Inflation

… **on** …

- Product sales
- Store footfall
- This customers' spend at this retailer



e.g. 'What is the effect of advertising this ice cream on its sales?'

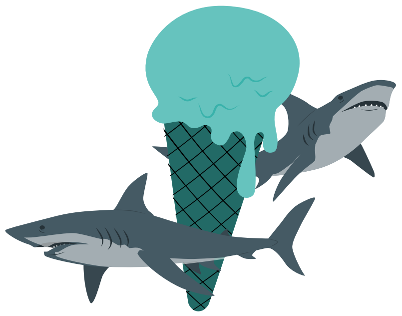
Or, broader questions:

- **Why** have this product's sales fallen?
- What are the **drivers** behind, or the factors affecting, product sales?

Other terms used in causal questions:

- **Incrementality, Uplift**
- **Test/Control** groups**, A/B tests, trials; matching**

From correlation to causation – what to adjust for?



Adjust for season

Adjust for income

e.g. 'What is the effect of advertising this product on its sales?'

What to adjust for, to isolate causal effect from spurious correlations?

Many possibly relevant variables, capturing traits of e.g.:

- Product (e.g. price, quality)
- Category (e.g. sales)
- Competing products
- Competing retailers
- Seasonality, weather
- The economy
- Consumer attitudes

If adjust for all / the wrong ones: bias.

→ Graphical Causal methods needed

Data science pipelines comparison

Associational *Prediction, $\hat{y}$*

(Define the problem)

↓

DATA.
Feature selection

↓

Apply models
e.g. regression, matching

↓

Validate models
e.g. train/test; k-fold CV; RMSE

↓

Explainability
feature importance

Data science pipelines comparison

Associational

Prediction, $\hat{y}$

- (Define the problem)
- **DATA.** Feature selection
- **Apply models** e.g. regression, matching
- **Validate models** e.g. train/test; k-fold CV; RMSE
- **Explainability** feature importance

Causal

Causal explanation, $\hat{\beta}$ as causal effects

- **Define the problem** X, Y, other causes
- **1. Draw DAG** Data generating process
- **2. Identify effect** Remove bias: spurious correl.
- **3. Estimate on DATA:** Apply models (same)
- **4. Validate models (same)** + causal: data vs {DAG, estimate}
- **Explainability - causal** Baked in

Theory basics:

The causal pipeline, step by step

1. Graphical causal models

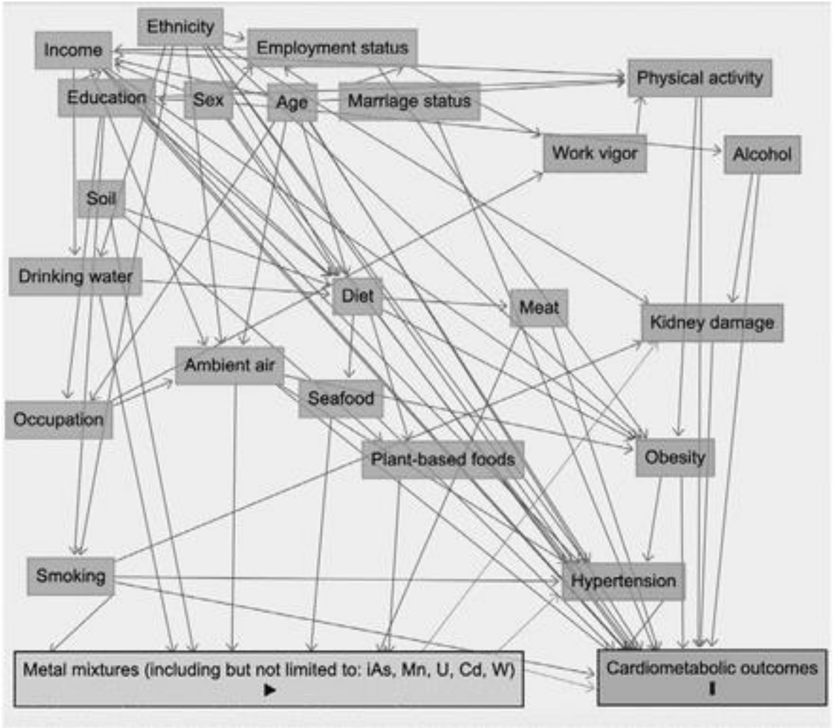
Causal diagrams, Causal Directed Acyclic Graphs (DAGs)

- **‘No causes in, no causes out’:** data alone isn't enough.
- Knowledge / assumptions about data generating process (DGP). E.g. domain knowledge, previous studies/ experiments.
- A → B: ‘A is *a possible* cause of B’.
- ***Absence*** of arrow from A to B: ‘A is **definitely not** a cause of B’. **Stronger assumption.** E.g. temporal precedence

→ ***Create DAG with team & stakeholders. Transparent.***

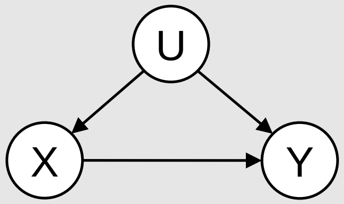
→ ***All possible: causal variables (even if not in data); arrows***

→ ***Complexity okay, reality is complex***



1. Graphical causal models - terminology

Causal diagrams, Causal Directed Acyclic Graphs (DAGs)



Join probability distribution

$P(y, x, u) = P(u)P(x|u)P(y|x, u)$

- In the DAG: A → B → C :

  A is a *parent* of B, B is a *child* of A.

  C is a *descendant* of A, A is an *ancestor* of C.

- Endogenous variable: has parent(s). Otherwise exogenous.
- A path*:* a sequence of arrows through variables, regardless of arrows' directions.
- Causal DAG: Mathematical object. Joint probability distribution. Causal Markov condition $\triangleq$ endogenous variables only depend on their parents. Can check vs. data

Some important path & variable types:

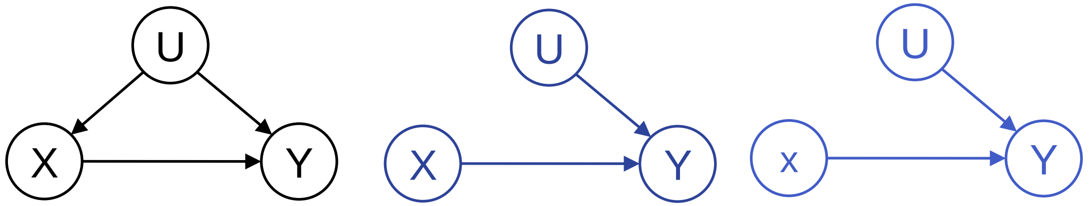
- A ← B → C. B: a **fork** on this path (***common cause***)
- Path A → B ← C. B: a **collider**

2. Identification: Isolating causal effects

- DAG structure → which variables to adjust for. Pure causal effect of X (the treatment) on Y (the outcome), free from spurious associations (bias)
- Adjusting for, controlling for **Z** = grouping by/ segregating by/ stratifying on **Z**
    - Regression: … on **Z** (input features: X and **Z**)
    - Matching: … on **Z**
- Adjusting for ALL variables can introduce bias (Berkson's paradox). Rather, use *just the right* variables.

Interventions

- Causal effects: **interventions, or surgeries, or manipulations**, on the DAG's arrows
- $P(Y=y \mid$ **do**$(X = x))$. 'The probability (or frequency) that event $Y= y$ would occur, if, hypothetically, X were set to the particular value x through **experimental manipulation** or **intervention.'**
- Performing a **surgery** on the DAG:
    - Delete all **arrows into** *X*,
    - Set *X*'s value to *x*, and leave the rest of the DAG unchanged.



Original model P

$P(y, x, u) = P(u)P(x|u)P(y|x, u)$

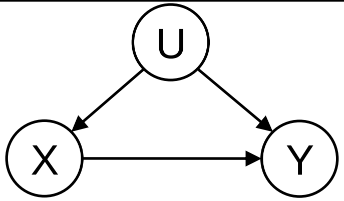
Manipulated model Pm

$Pm(y, x, u) = P(u)P(x)P(y|x, u)$

➢ This has resulted in the post-intervention DAG & distribution (Pm).

- Causal effect $P(Y=y|$ **do**$(X = x)) =$ **Pm**$(Y = y| X = x)$
- But in P: backdoor path via U. $P(Y=y \mid do(X=x)) \neq P(Y=y \mid X=x)$. → **Need identification step.**

2. Identification: Blocking or d-separation



A set of nodes ***Z*** block or d-separate a path *p* if & only if

i. p contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle node m is in set ***Z*** (i.e. m is conditioned on), or

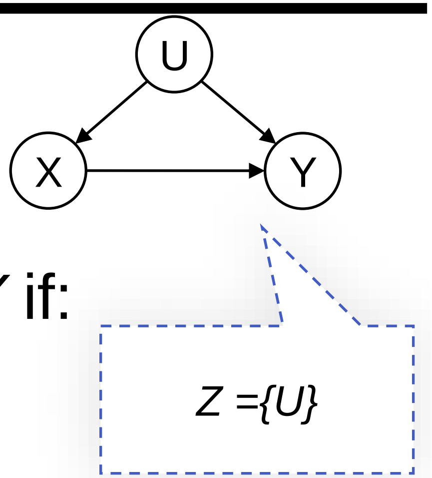
ii. p contains a collider $i \rightarrow \mathrm{m} \leftarrow j$ such that *neither* the collision node m, nor any of its descendants, are in set ***Z***.

Treat colliders differently! Colliders should NOT be in set Z!

If set ***Z*** blocks every path between two nodes *X* and *Y*, then X and Y are d-separated, conditional on **Z**, and thus are independent conditional on **Z**.

2. Identification: Backdoor criterion

Blocking all *backdoor (aka spurious)* paths.



A set of variables ***Z*** satisfies the backdoor criterion relative to *X* and *Y* if:

i. no node in ***Z*** is a descendant of *X*, and

ii. ***Z*** blocks every path between *X* and *Y* that contains an arrow into *X*.

Then ***Z***: a **sufficient, admissible or deconfounding set**.

The **minimal** deconfounding set: the smallest such ***Z*** that satisfies the backdoor criterion.

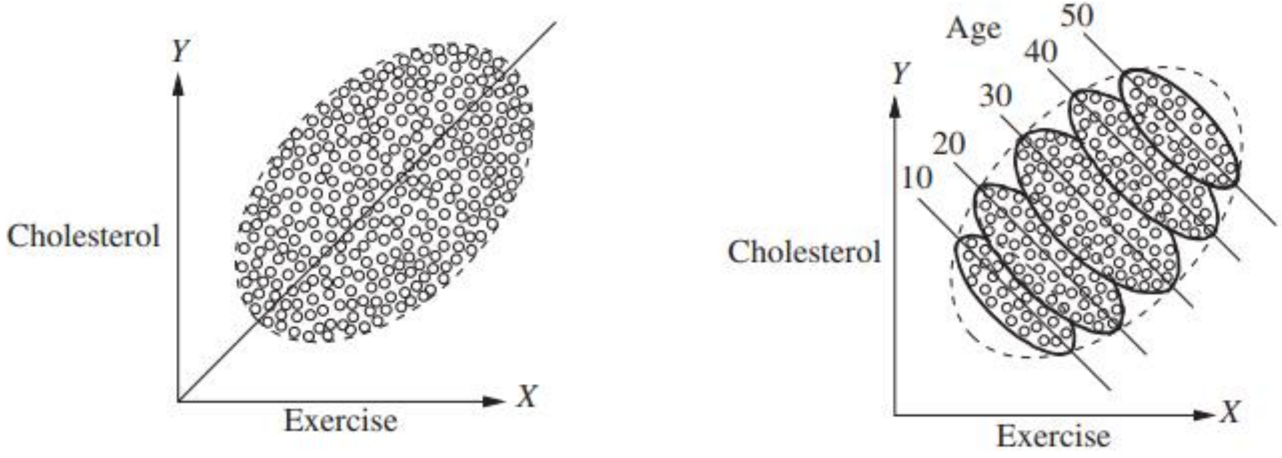
Adjust for deconfounding set **Z**: confounding bias (spurious associations) removed, causal effect *X* on *Y* identifiable from observational data.

Simpson's paradox

Different (even reversed) effect in whole population vs. in subpopulations

Example Question: What is the effect of weekly exercise on cholesterol level?

Opposite trends, depending on whether adjust for (group by) age or not!



→ Which one gives the right answer?

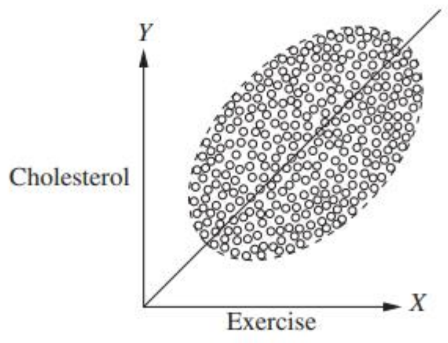
Source: Pearl, Glymour, Jewell 'Primer' book, p 3-4.

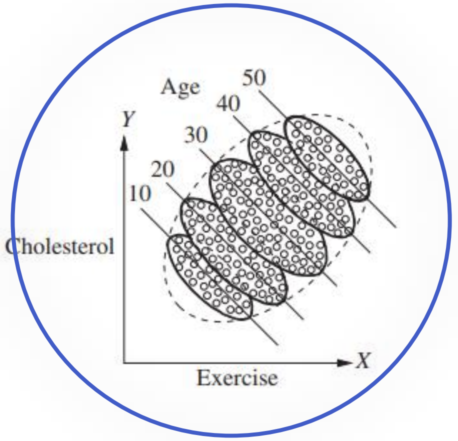
Simpson's paradox

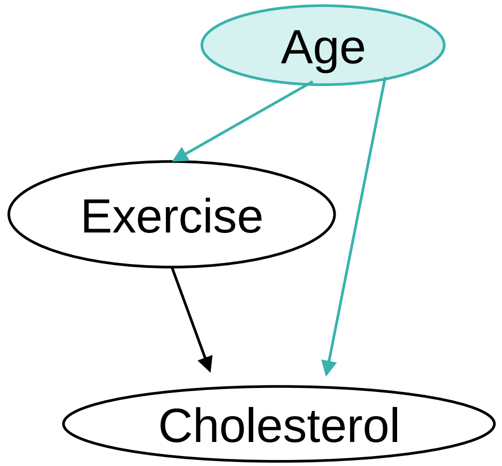
Different (even reversed) effect in whole population vs. in subpopulations

Example Question: What is the effect of weekly exercise on cholesterol level?

CAUSAL SOLUTION







- If know that age is a cause of exercise uptake, & of cholesterol levels, DAG:
- **Must** adjust for **confounders** (forks)
  - common causes of X & Y. Minimal set.

Source: Pearl, Glymrour, Jewell 'Primer' book, p 3-4.

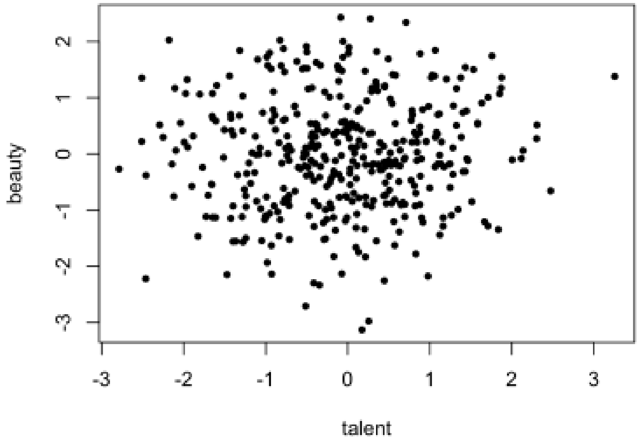
Berkson's paradox

E.g. False observation of negative correlation between two desirable traits

Example Question:

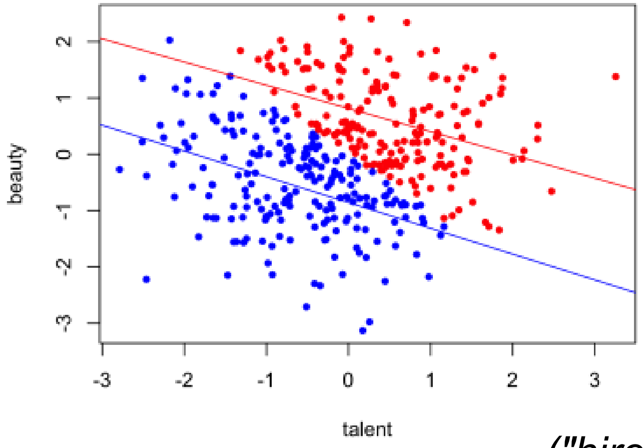
Are an actor's beauty & talent correlated?

No vs. negative relationship, depending on whether adjust for Hire status!



General Population

(no association)



Within Z=1 or Z=0 *("hire")*

(negative association)

→ Which one gives the right answer?

Source: Cinelli et al. 'Crash Course in Good & Bad controls' slides

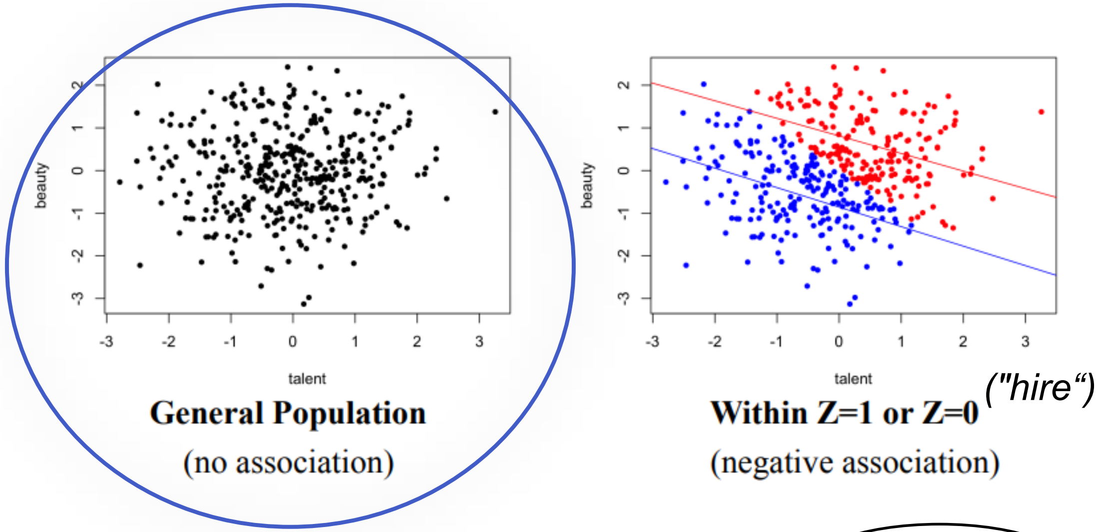
Berkson's paradox

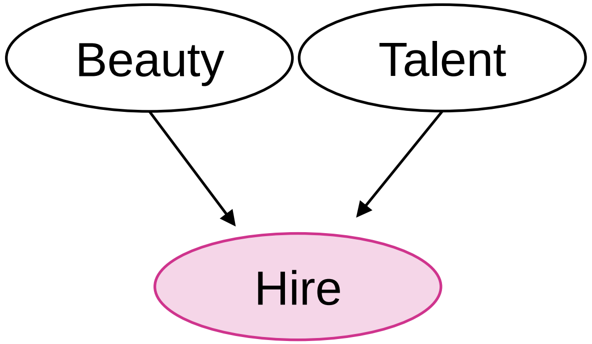
E.g. False observation of negative correlation between two desirable traits

Example Question:

Are an actor's beauty & talent correlated?

CAUSAL SOLUTION



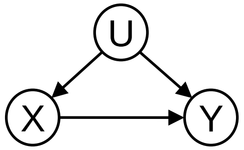


- If know that hire depends on beauty and talent, DAG:
- Must **NOT** adjust for **colliders** - common outcomes of X & Y or of *their* causes (M-bias)

Similarly for more paradoxes: Antebellum puzzle, Birth-weight paradox,…

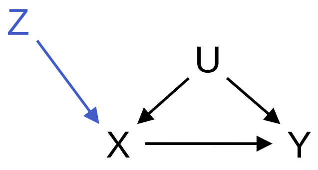
Source: Cinelli et al. 'Crash Course in Good & Bad controls' slides

2. Identification - Other estimands: IV, front-door

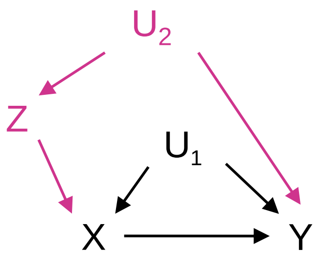


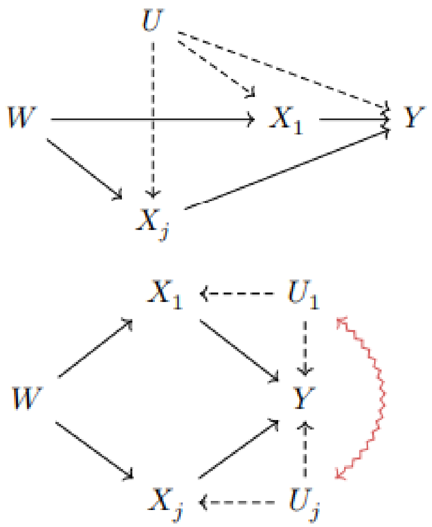
- Value: if *any* confounder(s) U between X & Y are *unmeasured*, still identify effect
- **Still assume: no *further* unblocked unmeasured confounding**

Instrumental variables (IV), aka *surrogate experiments*

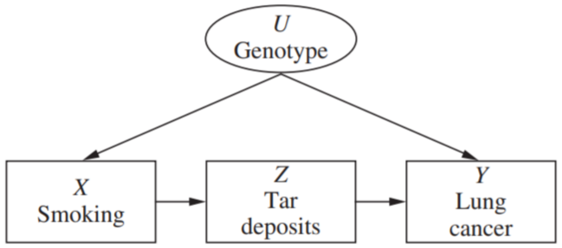


BUT:





Front-door criterion



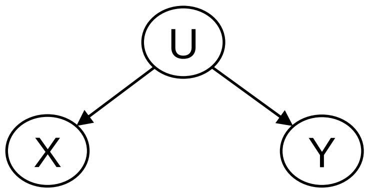
3. Estimation: Key points

- **Backdoor estimand:**
  - **Standard DS estimation models/methods** e.g. regression (linear or nonlinear); matching.
    - **Regression:**
      - **Linear**: coefficient now causal effect. (if many/ all Xs: run identification on all these).
      - **Nonlinear: partial dependence plots, sklearn:** Decision Trees, Random Forests, Gradient Boosting. Not ‘variable importance’ (associational, contrib. to goodness of fit).
    - Other newer **estimation** methods: e.g. Meta-Learners; Double ML; Doubly Robust Learning; synthetic controls; propensity score matching. **Assume unconfoundedness/ must still 1. Draw DAG and 2. Identify - deconfound**
- ***If found valid IV or front-door estimand.*** *E.g. if linear system: Wald estimator, 2SLS; 2SLS.*

4. Validation: Key points

- Causal DAGs have testable implications in the data sets they generate
- Know **that** the model is wrong, & **where** it is wrong. Standard stat. tests.

- **D-separation in DAG - conditional independence in data**
  If **Z** d-separates X from Y, then $(X \perp\!\!\!\perp Y \mid \mathbf{Z})$ ($\equiv P(Y|X, \mathbf{Z}) = P(Y|\mathbf{Z})$),



DAG: $X \perp\!\!\!\perp Y \mid U$

→ in data: should find no significant dependence between X & Y cond. on U.

→ Else: DAG inconsistent with DGP of dataset. E.g. maybe other path from X to Y, or from Y to X, or other common causes between X and Y.

- + further methods for **refuting** the final **estimate** (DoWhy).

Data science pipelines comparison

Associational

Prediction, $\hat{y}$

(Define the problem)

DATA.
Feature selection

Apply models
e.g. regression, matching

Validate models
e.g. train/test; k-fold CV; RMSE

Explainability
feature importance

Causal

Causal explanation,
$\hat{\beta}$ as causal effects

Define the problem
X, Y, other causes

1. Draw DAG
Data generating process

New

2. Identify effect
What should/ not adjust for

New *but* ***automated***

3. Estimate on DATA:
Apply models (same)

Same

4. Validate models (same)
+ causal: data vs {DAG, estimate}

Same *+* ***new*** *but mostly* ***automated***

Explainability - causal
Coeffs / PDP

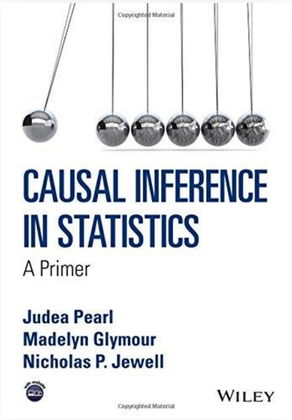
Baked in *stronger interp.,* ***causal***

Transparent, Explainable, Reliable AI

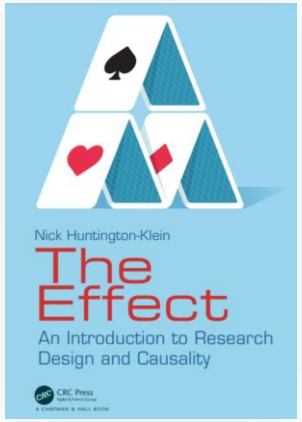
If you'd like to learn more...

Carlos Cinelli Andrew Forney, and Judea Pearl (2022+). "A Crash Course in Good and Bad Controls." **Sociological Methods and Research.**
[abstract] [preprint]
[journal] [slides]
[r code] [python code]

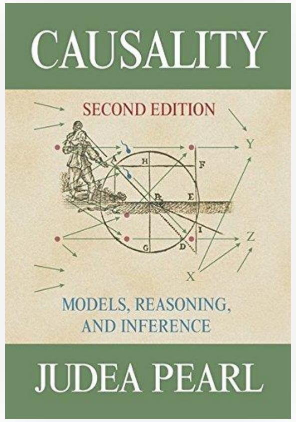
Paper, free online, DAGs, Python code



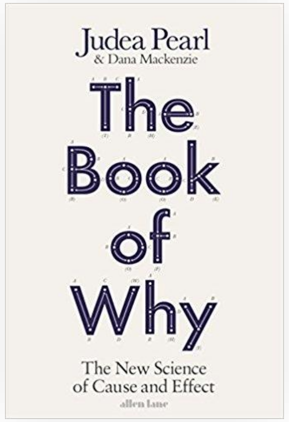
Mostly free online



All free online. Python & R code examples



More in-depth textbook



More general audience. Many parts free online

Practice: Applying the causal pipeline to a retail problem

Free open-source causal inference tools

DoWhy

DAGitty

Thank you

dunnhumby