Machine Learning 2.15: Causality

Tom S. F. Haines
T.S.F.Haines@bath.ac.uk

UNIVERSITY OF BATH

A fox eats the day

+

= annoying noise

UNIVERSITY OF BATH

A fox eats the day

+

= annoying noise

- If a fox eats the rooster, does the sun still rise?

UNIVERSITY OF BATH

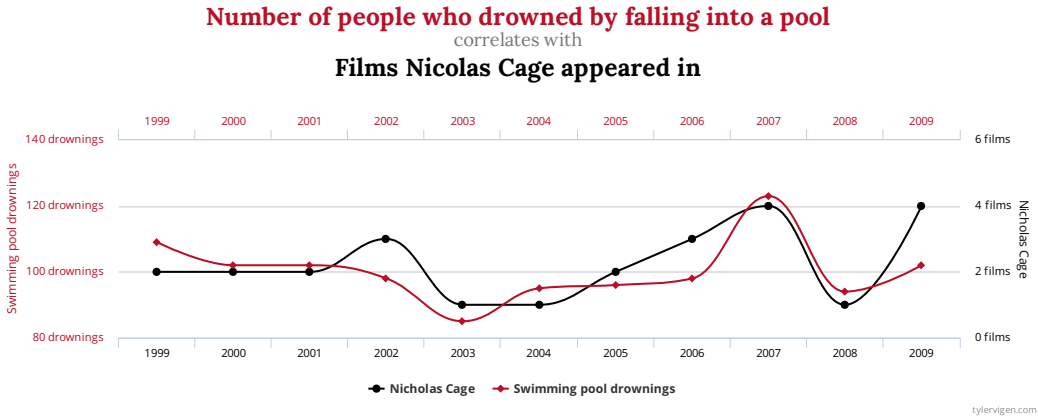
A fox eats the day

+

= annoying noise

- If a fox eats the rooster, does the sun still rise?
- Obviously yes
- **Correlation** does not know either way!

Spurious correlation



- **Machine learning = correlation**
  (exceptions in reinforcement learning)

- Suffers spurious correlation
- OK for prediction, terrible for *everything else*

UNIVERSITY OF BATH

Wrong correlation

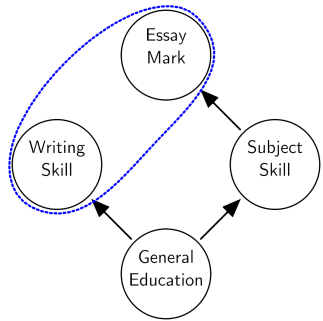
ML failures from lecture 4 of ML 1

- Perfect M-48 tank finding algorithm (1964)
- Tank photos taken on cloudy day
  Dark photo $\implies$ Tank
- Not-tank photos taken on sunny day
  Bright photo $\implies$ No tank

Wrong correlation

ML failures from lecture 4 of ML 1

- Perfect M-48 tank finding algorithm (1964)
- Tank photos taken on cloudy day
  Dark photo $\implies$ Tank
- Not-tank photos taken on sunny day
  Bright photo $\implies$ No tank



- Automatic essay marker
- Good writing $\implies$ High mark

Causality

- Problem: Machine learning = correlation
- Solution: Introduce **causality**!
  - "*Sun* **causes** *rooster to crow*"
  - Animals (inc. humans) do this

UNIVERSITY OF BATH

Causality

- Problem: Machine learning = correlation
- Solution: Introduce **causality**!
  - "*Sun* **causes** *rooster to crow*"
  - Animals (inc. humans) do this
- This lecture:
  - Introduction to causality
  - Some usable solutions
  - Limits

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

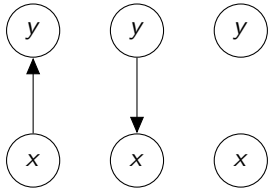
- You've all heard this...
  ...but what does it mean?

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



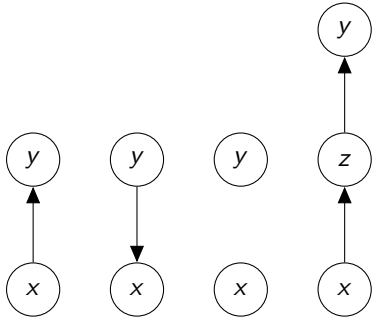
x correlated with y?	✓	✓	×
x causes y	✓	×	×

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



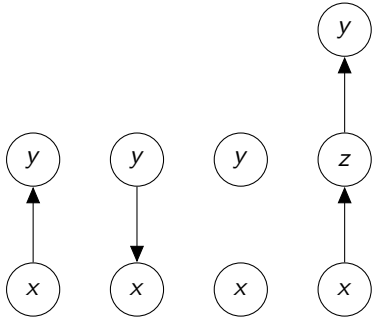
x correlated with y?	✓	✓	×
x causes y	✓	×	×

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



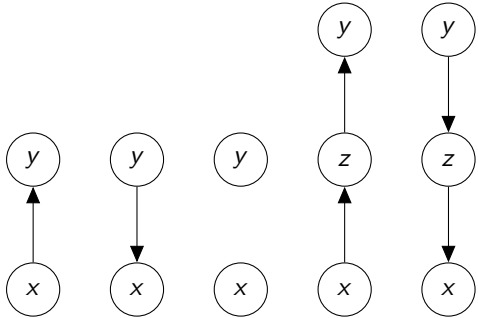
x correlated with y?	✓	✓	×	✓
x causes y	✓	×	×	✓

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



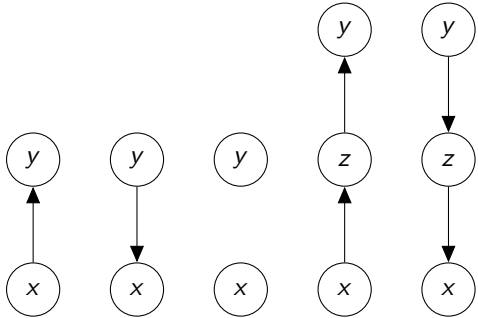
x correlated with y?	✓	✓	×	✓
x causes y	✓	×	×	✓

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



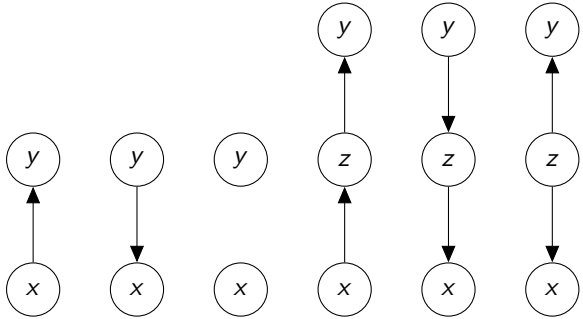
x correlated with y?	✓	✓	×	✓	✓
x causes y	✓	×	×	✓	×

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



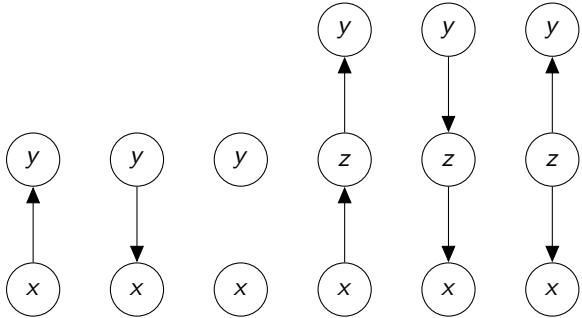
x correlated with y?	✓	✓	×	✓	✓
x causes y	✓	×	×	✓	×

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



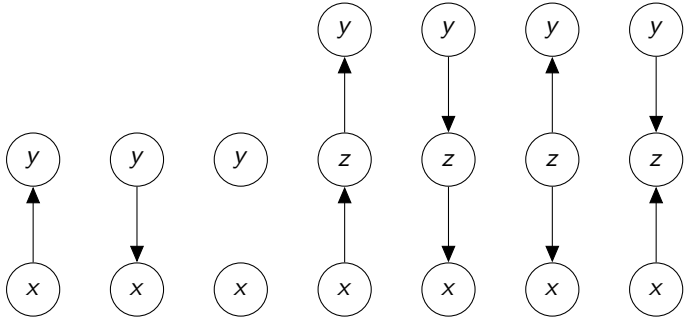
x correlated with y?	✓	✓	×	✓	✓	✓
x causes y	✓	×	×	✓	×	×

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



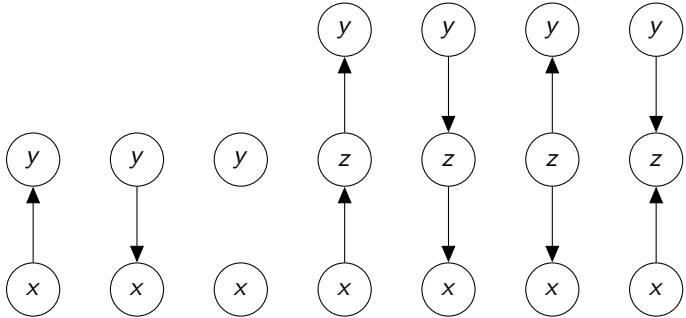
x correlated with y?	✓	✓	×	✓	✓	✓
x causes y	✓	×	×	✓	×	×

UNIVERSITY OF BATH

$\neq$

Correlation is not causation

- You've all heard this...
  ...but what does it mean?



x correlated with y?	✓	✓	×	✓	✓	✓	×
x causes y	✓	×	×	✓	×	×	×

Pearson correlation coefficient

Of greater use:

No correlation $\implies$ **no causation**

UNIVERSITY OF BATH

Pearson correlation coefficient

Of greater use:

No correlation $\implies$ no causation

- Usually Pearson correlation coefficient
- Linear correlation only

$$\begin{aligned}\rho_{X,Y} &= \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} \\ &= \frac{\mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right]}{\sqrt{\mathbb{E}\left[(X-\mathbb{E}[X])^2\right]}\sqrt{\mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right]}}\end{aligned}$$

Pearson correlation coefficient

Of greater use:

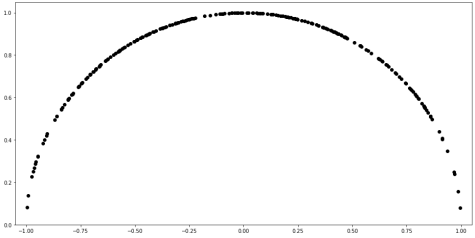
No correlation $\implies$ no causation

- Usually Pearson correlation coefficient
- Linear correlation only

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]}{\sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}\sqrt{\mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right]}}$$

- Non-linear correlation can "trick" it:

$$y = \sin(\cos^{-1}(x)), \quad x \sim \text{Uniform}(-1, 1)$$



Pearson correlation coefficient

Of greater use:

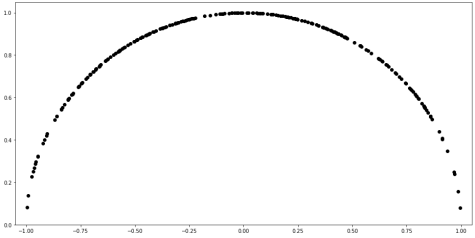
No correlation $\implies$ no causation

- Usually Pearson correlation coefficient
- Linear correlation only

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]}{\sqrt{\mathbb{E}\left[(X - \mathbb{E}[X])^2\right]}\sqrt{\mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right]}}$$

- Non-linear correlation can "trick" it:

$$y = \sin(\cos^{-1}(x)), \quad x \sim \text{Uniform}(-1, 1)$$



- No correlation $\implies$ no causation: Probably, but not perfectly
- ML can be better (learn $y = f(x)$; no better than random *implies* not correlated)

Bradford Hill criteria

- "**Does smoking cause cancer?**"
  - 1950: Statistics couldn't decide!
    i.e. could argue either way
    (no causal tools)

- "**Does smoking cause cancer?**"
  - 1950: Statistics couldn't decide!
    i.e. could argue either way
    (no causal tools)
  - Decision ultimately made in 1964
    by US surgeon general
  - Used "informal guidelines" of
    Austin Bradford Hill (published 1965)

Bradford Hill criteria

1. Strength (effect size, usually correlation)
2. Consistency
   (reproducible under variety of circumstances)
3. Specificity
   (it's precise, other explanations unlikelly)
4. Temporality (no time travel)
5. Biological gradient
   (greater exposure = stronger effect)
6. Plausibility (a sensible explanation!)
7. Coherence
   (lab experiments match observations)
8. Experiment (conduct one!)
9. Analogy (similar related causal patterns)
10. Reversibility
    (stopping cause disables effect; optional)

UNIVERSITY OF BATH

- "**Does smoking cause cancer?**"
  - 1950: Statistics couldn't decide!
    i.e. could argue either way
    (no causal tools)
  - Decision ultimately made in 1964
    by US surgeon general
  - Used "informal guidelines" of
    Austin Bradford Hill (published 1965)
  - Qualitative, fuzzy, subjective etc.
  - **Not maths!**
    – can't turn into an algorithm
  - Still used, esp. in epidemiology

Bradford Hill criteria

1. Strength (effect size, usually correlation)
2. Consistency
   (reproducible under variety of circumstances)
3. Specificity
   (it's precise, other explanations unlikelly)
4. Temporality (no time travel)
5. Biological gradient
   (greater exposure = stronger effect)
6. Plausibility (a sensible explanation!)
7. Coherence
   (lab experiments match observations)
8. Experiment (conduct one!)
9. Analogy (similar related causal patterns)
10. Reversibility
    (stopping cause disables effect; optional)

Ladder of causation

- From Judea Pearl's "*The Book of Why*"
- Three levels of reasoning
  (question kinds your model can answer)

Ladder of causation

- From Judea Pearl's "*The Book of Why*"
- Three levels of reasoning
  (question kinds your model can answer)

1. Association / seeing:
   - Given these symptoms which disease is most likely?
   - Will you like film x given you like film y?

   (you can already do this with ML)

Ladder of causation

- From Judea Pearl's "*The Book of Why*"
- Three levels of reasoning
  (question kinds your model can answer)

1. Association / seeing:
   - Given these symptoms which disease is most likely?
   - Will you like film x given you like film y?

   (you can already do this with ML)

2. Intervention / doing:
   - What if I press this big red button?
   - Does this pill cure cancer?

Ladder of causation

- From Judea Pearl's "*The Book of Why*"
- Three levels of reasoning
  (question kinds your model can answer)

1. Association / seeing:
   - Given these symptoms which disease is most likely?
   - Will you like film x given you like film y?

   (you can already do this with ML)

2. Intervention / doing:
   - What if I press this big red button?
   - Does this pill cure cancer?

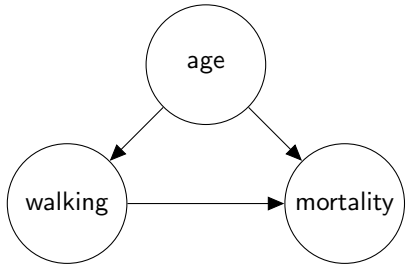
3. Counterfactuals / imagining:
   - Did immunisation prevent me from dying?
   - What if the moon was made of cheese?

Causal graphs

- Bayesian network (from lecture 11, ML 1)
- Introduces structure (factorisation)

$$P(\text{walking}, \text{age}, \text{mortality}) =$$

$$P(\text{mortality}|\text{walking}, \text{age})P(\text{walking}|\text{age})P(\text{age})$$

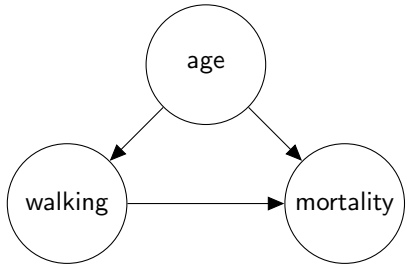


Causal graphs

- Bayesian network (from lecture 11, ML 1)
- Introduces structure (factorisation)

$$P(\text{walking}, \text{age}, \text{mortality}) =$$

$$P(\text{mortality}|\text{walking}, \text{age})P(\text{walking}|\text{age})P(\text{age})$$



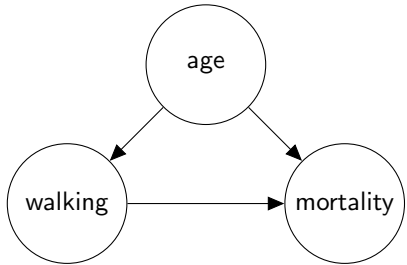
- $x \to y$ means x *causes* y
- Causal interpretation previously optional; required for *causal graph*

Causal graphs

- Bayesian network (from lecture 11, ML 1)
- Introduces structure (factorisation)

$$P(\text{walking}, \text{age}, \text{mortality}) =$$

$$P(\text{mortality}|\text{walking}, \text{age})P(\text{walking}|\text{age})P(\text{age})$$

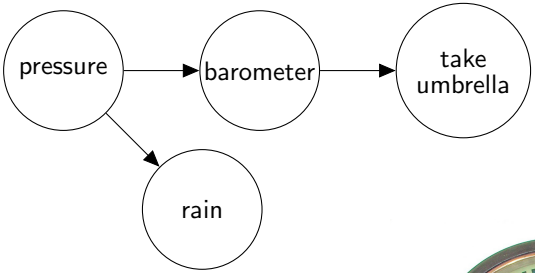


- $x \to y$ means x *causes* y
- Causal interpretation previously optional; required for *causal graph*

- Rung 1 (seeing) questions:
  Calculating conditional probabilities
- Bayes net does this (belief propagation etc.)

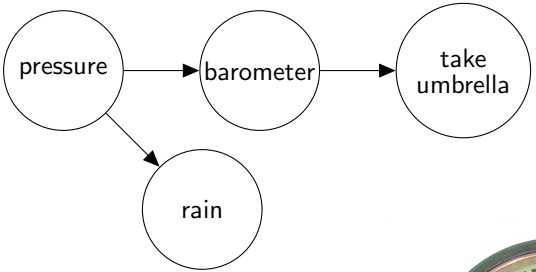
Action

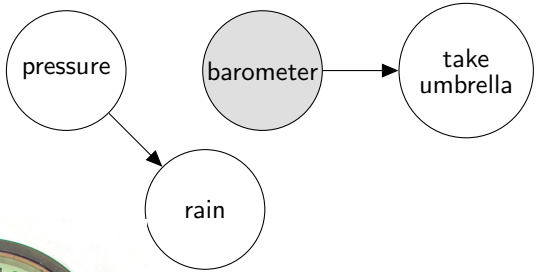
- Rung 2 (doing): What happens if you move the needle?



Action

- Rung 2 (doing): What happens if you move the needle?





$\mathrm{do}(\cdot)$

- Action **changes** the model!
- Not a conditional probability

do(·)

- Action **changes** the model!
- Not a conditional probability

- New operator:

$$P\left(\text{take umbrella}|\,\text{do}(\text{barometer} = \text{rain})\right)$$

- Meaning:
  1. **Delete** arrows pointing at RVs affected by do(·)
  2. Fix value of RVs affected by do(·)
  3. Calculate conditional probability for **new model**

do(·)

- Action **changes** the model!
- Not a conditional probability

- New operator:

$$P\left(\text{take umbrella}|\,\text{do}(\text{barometer} = \text{rain})\right)$$

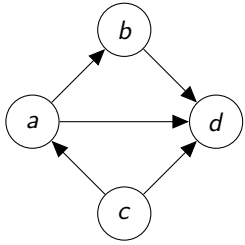
- Meaning:
  1. **Delete** arrows pointing at RVs affected by do(·)
  2. Fix value of RVs affected by do(·)
  3. Calculate conditional probability for **new model**

- Can do(·) several RVs and observe variables at same time, e.g.

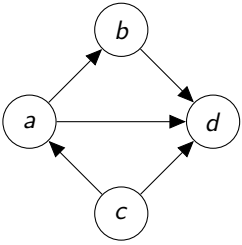
$$P\left(y|x, \text{do}(z), \text{do}(a)\right)$$

UNIVERSITY OF BATH

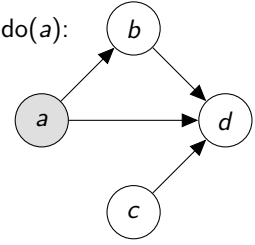
Example



Example

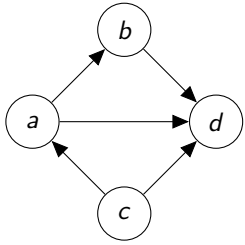


do(a):

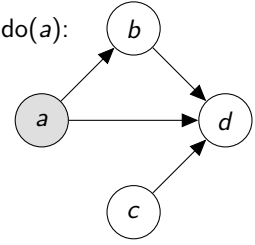


UNIVERSITY OF BATH

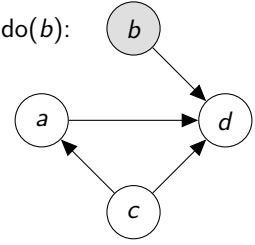
Example



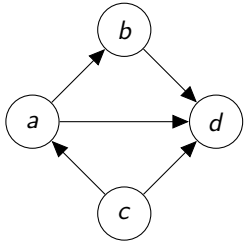
do(a):



do(b):



Example

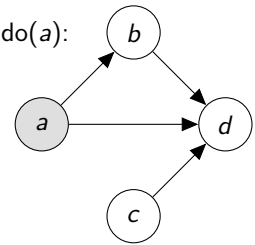


- For $\mathrm{do}(c)$ model unchanged $\therefore$

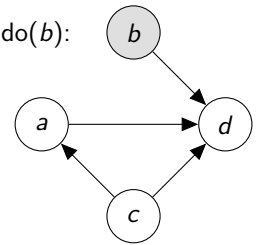
$$P(\cdot \mid \mathrm{do}(c)) = P(\cdot \mid c)$$

(assuming model correct,
e.g. no unknown RVs)

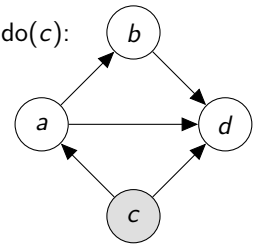
$\mathrm{do}(a)$:



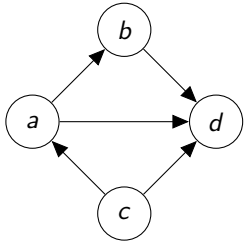
$\mathrm{do}(b)$:



$\mathrm{do}(c)$:



Example

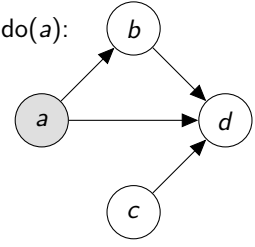


- For $\mathrm{do}(c)$ model unchanged $\therefore$

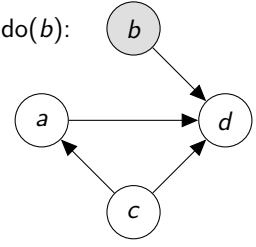
$$P(\cdot \mid \mathrm{do}(c)) = P(\cdot \mid c)$$

(assuming model correct,
e.g. no unknown RVs)

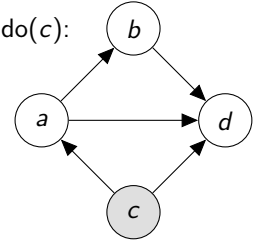
$\mathrm{do}(a)$:



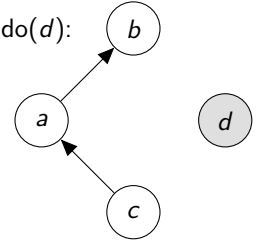
$\mathrm{do}(b)$:



$\mathrm{do}(c)$:



$\mathrm{do}(d)$:



Randomised controlled trial

RCT

- Typical medical scenario:
  1. Split guinea pigs into two groups **at random**
  2. Give half drug, half placebo (control group)
  3. Measure performance
  4. Difference between groups = drug performance

Randomised controlled trial

RCT

- Typical medical scenario:
  1. Split guinea pigs into two groups **at random**
  2. Give half drug, half placebo (control group)
  3. Measure performance
  4. Difference between groups = drug performance
- This is the do$(\cdot)$ operator!
  (randomly setting RV = deleting incoming arrows)
- It's causality using "*normal*" statistics

Randomised controlled trial

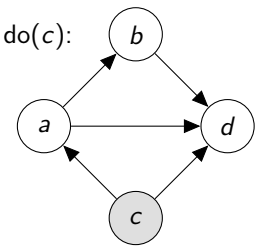
RCT

- Typical medical scenario:
  1. Split guinea pigs into two groups **at random**
  2. Give half drug, half placebo (control group)
  3. Measure performance
  4. Difference between groups = drug performance
- This is the do(·) operator!
  (randomly setting RV = deleting incoming arrows)
- It's causality using "*normal*" statistics
- RCTs: Hard, expensive, ethical limits
- Historically seen as **only** approach
- do(·) operator provides other approaches...

Observational studies

- *Observational study* or *natural experiment*
- *Action* from *observation*

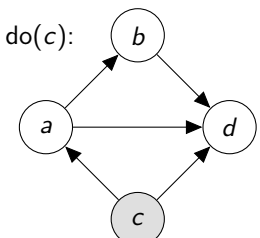
$$P(y|\,\mathrm{do}(x)) = P(y|x)$$



Observational studies

- *Observational study* or *natural experiment*
- *Action* from *observation*

$$P(y|\,\mathrm{do}(x)) = P(y|x)$$

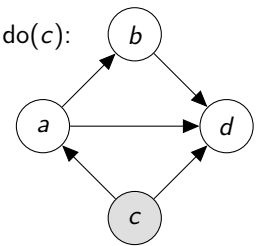


- Mendelian randomisation: Genetics as source of randomisation
  $P(y|\,\mathrm{do}(\mathrm{gene})) = P(y|\mathrm{gene})$ (assumes subjects don't know their genes!)

Observational studies

- *Observational study* or *natural experiment*
- *Action* from *observation*

$$P(y|\,\mathrm{do}(x)) = P(y|x)$$

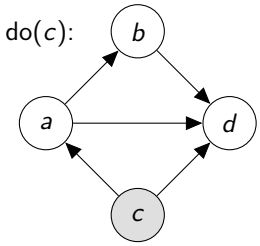


- Mendelian randomisation: Genetics as source of randomisation
  $P(y|\,\mathrm{do}(\mathrm{gene})) = P(y|\mathrm{gene})$ (assumes subjects don't know their genes!)
- e.g. Does smoking cause lung cancer?
  - Gene that increases nicotine addiction $\rightarrow$ smoke more (has no effect on non-smokers; know strength)
  - Compare lung cancer incidence between groups
  - Can calculate $P(\mathrm{lung\ cancer}|\,\mathrm{do}(\mathrm{smoke}))$

Observational studies

- *Observational study* or *natural experiment*
- *Action* from *observation*

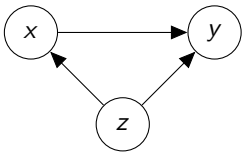
$$P(y|\,\mathrm{do}(x)) = P(y|x)$$



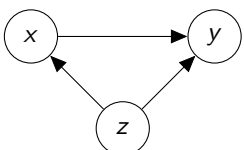
- Mendelian randomisation: Genetics as source of randomisation
  $P(y|\,\mathrm{do}(\mathrm{gene})) = P(y|\mathrm{gene})$ (assumes subjects don't know their genes!)
- e.g. Does smoking cause lung cancer?
  - Gene that increases nicotine addiction $\rightarrow$ smoke more (has no effect on non-smokers; know strength)
  - Compare lung cancer incidence between groups
  - Can calculate $P(\mathrm{lung\ cancer}|\,\mathrm{do}(\mathrm{smoke}))$
- Limited: Can only do($\cdot$) RVs with no incoming arrows
- More general form...

Confounding

- Confounding RV (z):
  Common cause of independent (x) and dependent (y) RV
- Referred to as "*backdoor paths*"



Confounding



- Confounding RV (z):
  Common cause of independent (x) and dependent (y) RV
- Referred to as "*backdoor paths*"

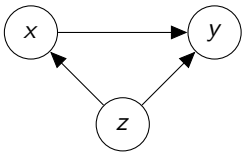
- Confounding breaks observational study:

$$P(y|\,\mathrm{do}(x)) \neq P(y|x)$$

- Fix by including confounding variables:

$$P(y, z|\,\mathrm{do}(x)) = P(y, z|x)$$

Confounding



- Confounding RV (z):
  Common cause of independent (x) and dependent (y) RV
- Referred to as "*backdoor paths*"

- Confounding breaks observational study:

$$P(y|\,\mathrm{do}(x)) \neq P(y|x)$$

- Fix by including confounding variables:

$$P(y, z|\,\mathrm{do}(x)) = P(y, z|x)$$

- d-separation rules: Identify z for any graph...
  (from lecture 11 of ML 1)

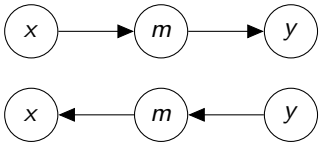
d-separation

- d = directional; opposite is d-connected
- **RVs x and y are d-separated if independent given observed RVs, z**
- Must consider all paths between x and y, ignoring directions

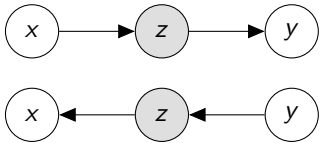
d-separation

- d = directional; opposite is d-connected
- **RVs x and y are d-separated if independent given observed RVs, z**
- Must consider all paths between x and y, ignoring directions

d-connected (dependent):



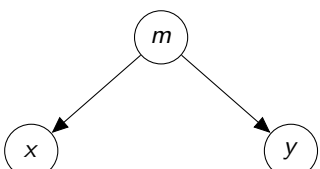
d-seperated (conditionally independent):



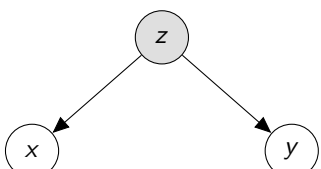
d-separation

- d = directional; opposite is d-connected
- **RVs x and y are d-separated if independent given observed RVs, z**
- Must consider all paths between x and y, ignoring directions

d-connected (dependent):



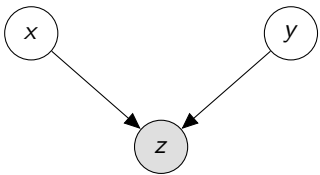
d-seperated (conditionally independent):



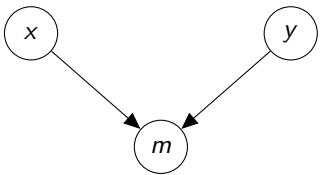
d-separation

- d = directional; opposite is d-connected
- **RVs x and y are d-separated if independent given observed RVs, z**
- Must consider all paths between x and y, ignoring directions

d-connected (dependent):



d-seperated (conditionally independent):

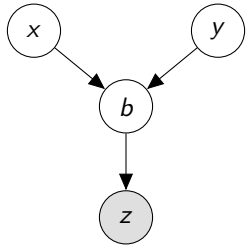


This is backwards!

d-separation

- d = directional; opposite is d-connected
- **RVs x and y are d-separated if independent given observed RVs, z**
- Must consider all paths between x and y, ignoring directions

d-connected (dependent):



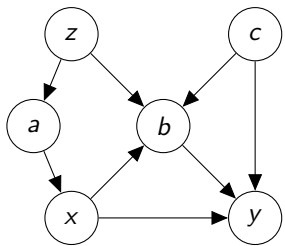
Can think in terms of information "bouncing off" of observed RVs (arrow heads only)

UNIVERSITY OF BATH

Closing the backdoor

- Don't want d-separation (conditional independence)
- Goal: Close *backdoors*, **without** blocking "*front doors*"

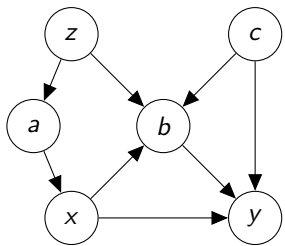
Goal: $P(y \mid \mathrm{do}(x))$



Closing the backdoor

- Don't want d-separation (conditional independence)
- Goal: Close *backdoors*, **without** blocking "*front doors*"

Goal: $P(y \mid \mathrm{do}(x))$

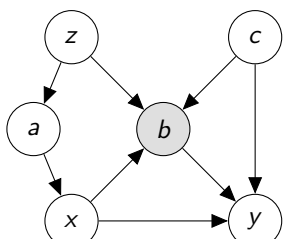


- Don't control – **invalid**
  - z confounds

Closing the backdoor

- Don't want d-separation (conditional independence)
- Goal: Close *backdoors*, **without** blocking "*front doors*"

Goal: $P(y \mid \mathrm{do}(x))$

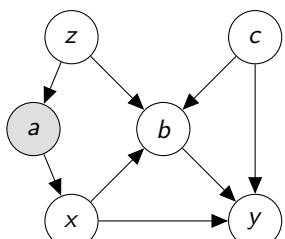


- Don't control – invalid
  - z confounds
- Control b – **invalid**
  - z and c confound – bounce off b
  - $x \to b \to y$ front door blocked

Closing the backdoor

- Don't want d-separation (conditional independence)
- Goal: Close *backdoors*, **without** blocking "*front doors*"

Goal: $P(y \mid \mathrm{do}(x))$

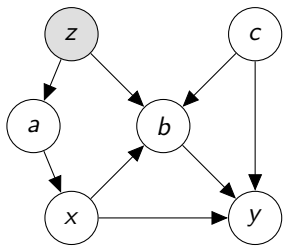


- Don't control – invalid
  - z confounds
- Control b – invalid
  - z and c confound – bounce off b
  - $x \to b \to y$ front door blocked
- Control a – **valid**

Closing the backdoor

- Don't want d-separation (conditional independence)
- Goal: Close *backdoors*, **without** blocking "*front doors*"

Goal: $P(y \mid \mathrm{do}(x))$

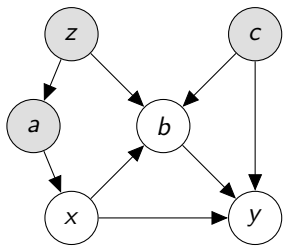


- Don't control – invalid
  - z confounds
- Control b – invalid
  - z and c confound – bounce off b
  - $x \to b \to y$ front door blocked
- Control a – valid
- Control z – **valid**

Closing the backdoor

- Don't want d-separation (conditional independence)
- Goal: Close *backdoors*, **without** blocking "*front doors*"

Goal: $P(y \mid \mathrm{do}(x))$



- Don't control – invalid
  - z confounds
- Control b – invalid
  - z and c confound – bounce off b
  - $x \to b \to y$ front door blocked
- Control a – valid
- Control z – **valid**

- No harm in controlling more than a or z, as long as front doors remain open

Backdoor criterion

- Called: **backdoor criterion**
- Main approach to *observational studies*

Steps:

1. Obtain causal graphical model
2. Identify z, set of variables that
   - Block *backdoor*s
   - Leave *front door* open
3. Calculate:

$$P(y, z|\,\mathrm{do}(x)) = P(y, z|x)$$

4. Integrate (or sum) out z:

$$P(y|\,\mathrm{do}(x)) = \int P(y, z|x)P(z)dz$$

UNIVERSITY OF BATH

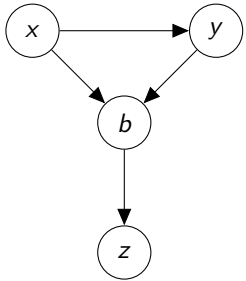
Over controlling

- Aside/warning
- Control wrong variable $\rightarrow$ create confounder!

UNIVERSITY OF BATH

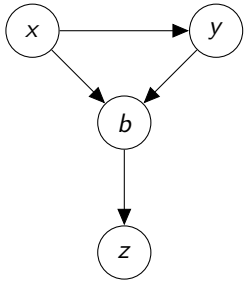
Over controlling

- Aside/warning
- Control wrong variable $\rightarrow$ create confounder!
- e.g. b or z



Over controlling

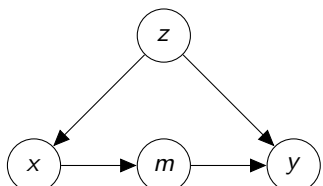
- Aside/warning
- Control wrong variable $\rightarrow$ create confounder!
- e.g. b or z
- Common mistake:
  Controlling for everything feels safe, but isn't
- Controlling reduces allocation bias in RCTs...
  ...so everything gets controlled – wrong!



Front door criterion

- Alternative: **front door criterion**
- Can't observe z $\therefore$ can't control it

Goal: $P(y|\,\mathrm{do}(x))$



Front door criterion

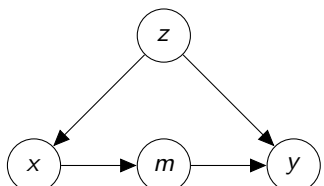
- Alternative: **front door criterion**
- Can't observe z $\therefore$ can't control it

- Can calculate: ($m =$ "*mediator*")

$$P(m|\,\mathrm{do}(x)) = P(m|x)$$

$$P(y|\,\mathrm{do}(m)) = P(y|m)$$

Goal: $P(y|\,\mathrm{do}(x))$



Front door criterion

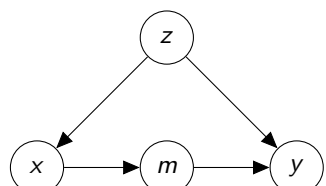
- Alternative: **front door criterion**
- Can't observe z $\therefore$ can't control it

- Can calculate: ($m =$ "*mediator*")

$$P(m|\,\mathrm{do}(x)) = P(m|x)$$

$$P(y|\,\mathrm{do}(m)) = P(y|m)$$

Goal: $P(y|\,\mathrm{do}(x))$



- "Concatenate":

$$P(y|\,\mathrm{do}(x)) = \int\int P(y|m, x = x')P(x = x')dx'P(m|x)dm$$

Front door criterion

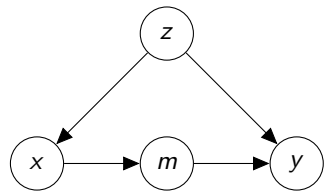
- Alternative: **front door criterion**
- Can't observe z $\therefore$ can't control it

- Can calculate: ($m =$ "*mediator*")

$$P(m|\,\mathrm{do}(x)) = P(m|x)$$

$$P(y|\,\mathrm{do}(m)) = P(y|m)$$

Goal: $P(y|\,\mathrm{do}(x))$



- "Concatenate":

$$P(y|\,\mathrm{do}(x)) = \int\int P(y|m, x = x')P(x = x')dx'P(m|x)dm$$

- Generalises: m blocks all front door paths
(without $x \to m$ or $m \to y$ having backdoors)

UNIVERSITY OF BATH

Rung 2 (doing) summary

- Action $= \text{do}(\cdot)$
  - Randomised controlled trial
  - Observational study

UNIVERSITY OF BATH

Rung 2 (doing) summary

- Action = do(·)
  - Randomised controlled trial
  - Observational study
- Approach depends on what you know
  - Ethical randomisation
  - Cost effective sample
  - Causal graphical model?
  - Observable variables vs unobservable variables
- Approaches can be impossible (sometimes all of them...)

UNIVERSITY OF BATH

Rung 2 (doing) summary

- Action $= \text{do}(\cdot)$
  - Randomised controlled trial
  - Observational study
- Approach depends on what you know
  - Ethical randomisation
  - Cost effective sample
  - Causal graphical model?
  - Observable variables vs unobservable variables
- Approaches can be impossible (sometimes all of them...)
- RCT safe if model incorrect – gold standard
  (other ways it can go wrong)
- Model errors break observational study
- Main risk: Unknown RVs

Do calculus

- Backdoor and front door derived from "*do calculus*"
- Can find problem specific solutions

Rule 1: If z blocks all paths from w to y:

$$P(y|z, w, \text{do}(x)) = P(y|z, \text{do}(x))$$

Rule 2: If z satisfies backdoor criterion:

$$P(y|\,\text{do}(x), z) = P(y|x, z)$$

Rule 3: If there is no front door path from x to y:

$$P(y|\,\text{do}(x)) = P(y)$$

+ probability rules

Rung 3 (imagining)

- All about **counterfactuals**
- Defined by David Hume: (1748)

"*if the first object had not been the second never had existed*"

Rung 3 (imagining)

- All about **counterfactuals**
- Defined by David Hume: (1748)

  "*if the first object had not been the second never had existed*"

- Reality (seen)

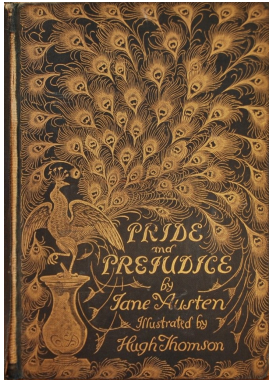
$$\text{Action} \rightarrow \text{Outcome}$$

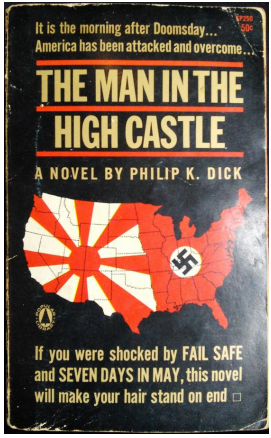
- Counterfactual (imagined)

$$\text{Counterfactual action} \rightarrow \text{Counterfactual outcome}$$

Entertainment / art

- You're familiar with this...





Perfect ninjas

- Can never observe counterfactuals!
- Requires assumptions – complete model
  - Causal graph
  - Distribution for each RV
- Wrong model $\rightarrow$ wrong counterfactual outcome

UNIVERSITY OF BATH

Example

- Goal: Predict salary for employee if they stayed longer in education

UNIVERSITY OF BATH

Example

- Goal: Predict salary for employee if they stayed longer in education
- Collect data: (example from *Book of Why*)

Employee	Experience (y)	Education (q)	Salary (s)
Alice	6	0	81
Bert	9	1	92.5
Caroline	9	2	97
David	8	1	91
Ernest	12	1	100
Frances	13	0	97

- Experience in years
- Education:
  - q=0 College
  - q=1 Bachelors degree
  - q=2 Masters degree
- Currency in thousands of pounds

Correlation

- Question is causal – correlation **does not work**
- But lets do it anyway!

Correlation

- Question is causal – correlation **does not work**
- But lets do it anyway!
- Treat as missing data problem

Employee	y	q	$s \mid \text{do}(q = 0)$	$s \mid \text{do}(q = 1)$	$s \mid \text{do}(q = 2)$
Alice	6	0	81	?	?
Bert	9	1	?	92.5	?
Caroline	9	2	?	?	97
David	8	1	?	91	?
Ernest	12	1	?	100	?
Frances	13	0	97	?	?

Correlation

- Question is causal – correlation **does not work**
- But lets do it anyway!
- Treat as missing data problem

Employee	y	q	$s\|\,\mathrm{do}(q=0)$	$s\|\,\mathrm{do}(q=1)$	$s\|\,\mathrm{do}(q=2)$
Alice	6	0	81	?	?
Bert	9	1	?	92.5	?
Caroline	9	2	?	?	97
David	8	1	?	91	?
Ernest	12	1	?	100	?
Frances	13	0	97	?	?

- Question: What if Bert had a masters?
- Nearest neighbours:
  Caroline has same number of years $\therefore$
  Bert: $s|\,\mathrm{do}(q=2) = 97$

Correlation

- Question is causal – correlation **does not work**
- But lets do it anyway!
- Treat as missing data problem

Employee	y	q	$s\|\,\mathrm{do}(q=0)$	$s\|\,\mathrm{do}(q=1)$	$s\|\,\mathrm{do}(q=2)$
Alice	6	0	81	?	?
Bert	9	1	?	92.5	?
Caroline	9	2	?	?	97
David	8	1	?	91	?
Ernest	12	1	?	100	?
Frances	13	0	97	?	?

- Question: What if Bert had a masters?
- Nearest neighbours:
  Caroline has same number of years $\therefore$
  Bert: $s|\,\mathrm{do}(q=2) = 97$

- What if Alice has an undergrad?
- Linear regression:

$$s = 65 + 2.5y + 5q$$

Alice: $s|\,\mathrm{do}(q=1) = 85$

UNIVERSITY OF BATH

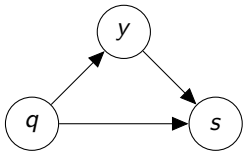
Structural causal model

- "*structural equation model*" (SEM) (linear)
- SEM invented by Sewell Wright
  (1921, probably first person to treat causality rigorously)
- "*structural causal model*" (SCM)
  (strictly causal, non-linear)

Structural causal model

- "*structural equation model*" (SEM) (linear)
- SEM invented by Sewell Wright
  (1921, probably first person to treat causality rigorously)
- "*structural causal model*" (SCM)
  (strictly causal, non-linear)

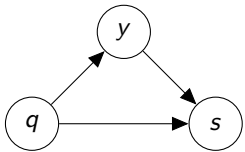
1. Causal graph
   ($q \to y$ because education reduces years of experience)



Structural causal model

- "*structural equation model*" (SEM) (linear)
- SEM invented by Sewell Wright
  (1921, probably first person to treat causality rigorously)
- "*structural causal model*" (SCM)
  (strictly causal, non-linear)

1. Causal graph
   ($q \to y$ because education reduces years of experience)
2. For each RV:
   - **Deterministic** function
   - Parametrised by causal RVs...
   - ...and noise parameters (N_*)
     (need distribution over each; can have many)



$$q = f_q(N_q)$$

$$y = f_y(q, N_y)$$

$$s = f_s(q, y, N_s)$$

(can be anything)

Counterfactual

- Train:
  1. Design structural causal model
  2. Fit free parameters to data (optimisation)

$$q = f_q(N_q) = N_q$$

$$y = f_y(q, N_y) = 10 - 4e + N_y$$

$$s = f_s(q, y, N_s) = 65 + 2.5y + 5q + N_s$$

Counterfactual

- Train:
  1. Design structural causal model
  2. Fit free parameters to data (optimisation)

$$q = f_q(N_q) = N_q$$

$$y = f_y(q, N_y) = 10 - 4e + N_y$$

$$s = f_s(q, y, N_s) = 65 + 2.5y + 5q + N_s$$

- Test:
  1. Fit noise parameters to individual (abduction)

- Fit individual (Alice):

$$81 = 65 + 2.5 \times 6 + 5 \times 0 + N_s \implies N_s = 1$$

$$6 = 10 - 4 \times 0 + N_y \implies N_y = -4$$

Counterfactual

- Train:
  1. Design structural causal model
  2. Fit free parameters to data (optimisation)

$$q = f_q(N_q) = N_q$$

$$y = f_y(q, N_y) = 10 - 4e + N_y$$

$$s = f_s(q, y, N_s) = 65 + 2.5y + 5q + N_s$$

- Test:
  1. Fit noise parameters to individual (abduction)
  2. Use do$(\cdot)$ to modify model (action)

- Fit individual (Alice):

$$81 = 65 + 2.5 \times 6 + 5 \times 0 + N_s \implies N_s = 1$$

$$6 = 10 - 4 \times 0 + N_y \implies N_y = -4$$

- do$(\cdot)$ operator:

$$s|\,\mathrm{do}(q) = 65 + 2.5(10 - 4q + N_y) + 5q + N_s$$

Counterfactual

- Train:
  1. Design structural causal model
  2. Fit free parameters to data (optimisation)

$$q = f_q(N_q) = N_q$$

$$y = f_y(q, N_y) = 10 - 4e + N_y$$

$$s = f_s(q, y, N_s) = 65 + 2.5y + 5q + N_s$$

- Test:
  1. Fit noise parameters to individual (abduction)
  2. Use do$(\cdot)$ to modify model (action)
  3. Evaluate desired RVs (prediction)

- Fit individual (Alice):

$$81 = 65 + 2.5 \times 6 + 5 \times 0 + N_s \implies N_s = 1$$

$$6 = 10 - 4 \times 0 + N_y \implies N_y = -4$$

- do$(\cdot)$ operator:

$$s|\,\mathrm{do}(q) = 65 + 2.5(10 - 4q + N_y) + 5q + N_s$$

- Predict:

$$76 = s|\,\mathrm{do}(q = 1) = 65 + 2.5(10 - 4 - 4) + 5 + 1$$

(linear regression was 85)

Summary

- Causality = asking different questions of your model

- Causal graphical models
- do(·) operator
- Structural causal models

- Scenarios with no solution exist
- A research area

- Where does causal model come from?
- Future of ML?

UNIVERSITY OF BATH

Further reading

- Complete guide:
  "*The Book of Why: The New Science of Cause and Effect*",
  by Pearl & Mackenzie (2018)
  (only real choice, but quite a lot of mistakes!)

UNIVERSITY OF BATH

Sources

- Some slides stolen from Julian Faraway
- Many examples from Judea Pearl
- Barrometer: CC Attribution 4.0 International, Auckland Museum
  `https://commons.wikimedia.org/wiki/File:Barometer,_aneroid_(AM_70061-7).jpg`
- Film posters and book covers – blatantly stolen