Causal inference : an introduction.

CentraleSupélec

Causality, Association

An **associational concept** is any relationship that can be defined in terms of a joint distribution of observed variables

A **causal concept** is any relationship that cannot be defined from the distribution alone.

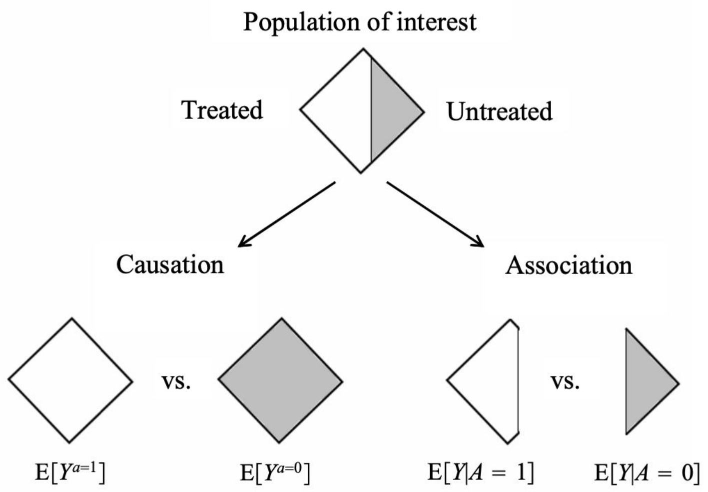
Examples of an associational concepts :

- correlation,
- regression

Examples of a causal concept :

- randomization,
- confounding

(Judea Pearl : the mathematics of causal relations)



Average Causal effect = $E(Y^{a=1}) - E(Y^{a=0})$

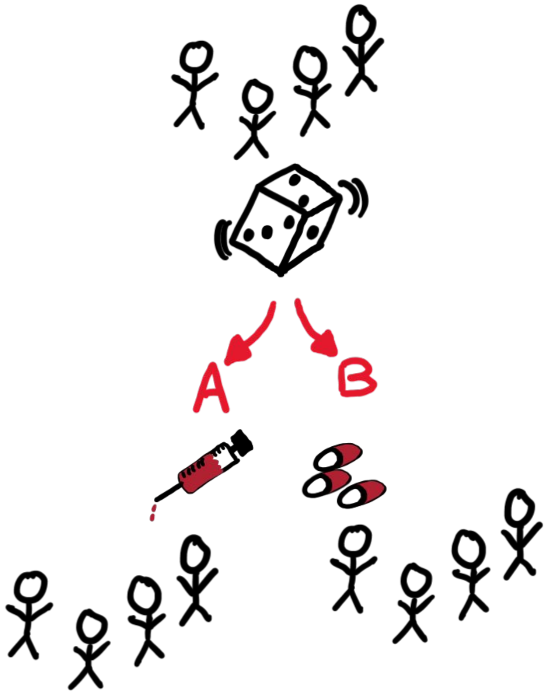
CentraleSupélec

Randomized experimentation, Observational studies and Bias

A randomized experiment is when the investigator carried out the action of interest and it was randomized because the decision to act on any study subject was made by a random device.
Randomization results in convincing causal inferences, the downsides being mostly financial and ethical.

A scientific study in which the investigator observes and records the relevant data is referred to as an observational study. And while observational studies are not considered robust causal inference tools,much human knowledge is derived from them.

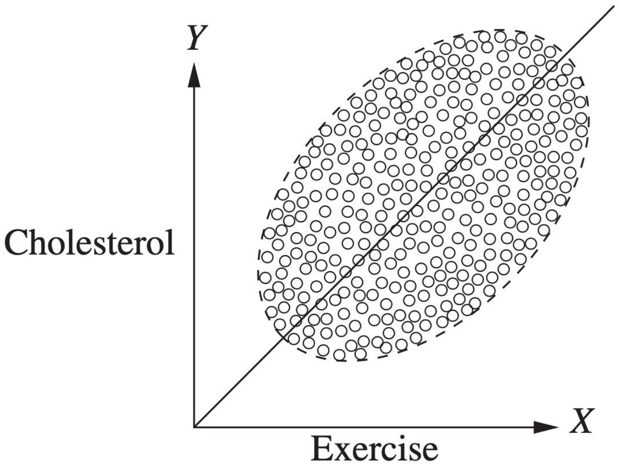
Statistical data, derived from studies and experimentation, can represent many forms of bias. The two most relevant to causal inference are Measurement Bias and Selection Bias.

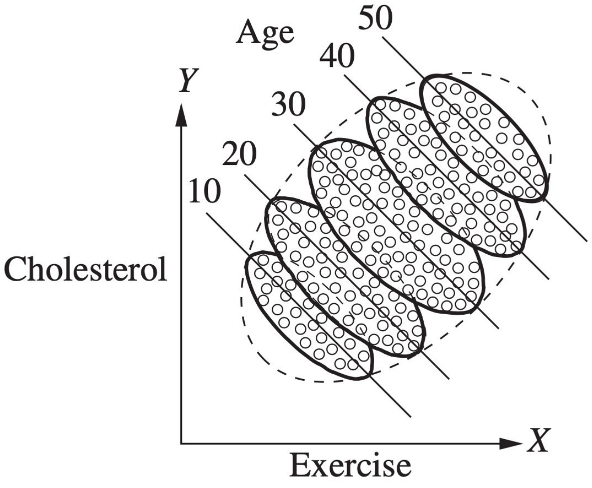


CentraleSupélec

Simpson’s Paradox

Simpson’s paradox, also called **Yule-Simpson effect**, in statistics, an effect that occurs when the marginal association between two categorical variables is qualitatively different from the partial association between the same two variables after controlling for one or more other variables.
(Britannica)



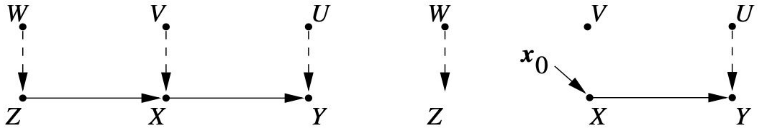


CentraleSupélec

Interactions, Counterfactuals

- Z, X, Y : endogenous variables
- W, V, U : exogenous variables, not influenced by the endogenous variables

Graph representation



(a) (b)

$$z = f_Z(w)$$

$$\begin{aligned} x &= f_X(z, v) \\ y &= f_Y(x, u) \end{aligned} \qquad \begin{aligned} x &= x_0 \\ y &= f_Y(x, u) \end{aligned}$$

Figure 1. The causal hierarchy. Questions at level 1 can be answered only if information from level i or higher is available.

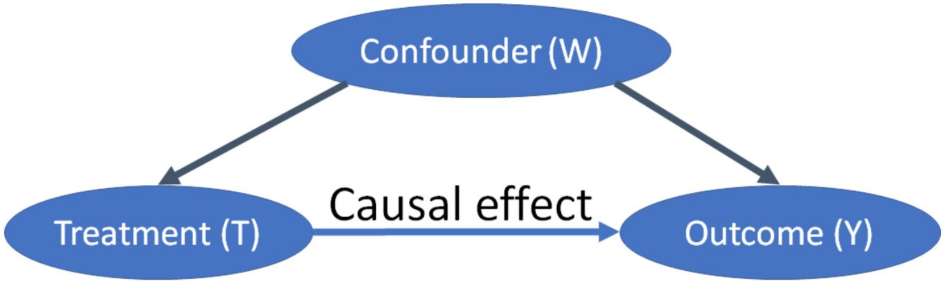
Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief inY?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing, Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past two years?

Judea Pearl : The Seven Tools of Causal Inference, with Reflections on Machine Learning

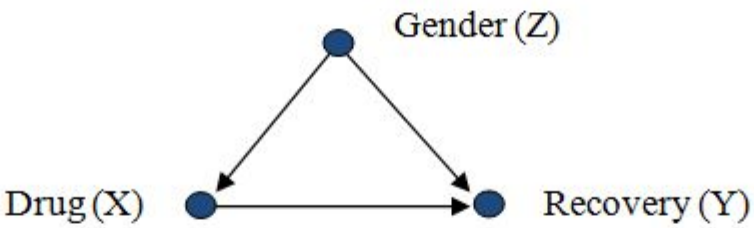
CentraleSupélec

Confounders

- variables that have effects on 2 variables leading to an association
- Causal concept



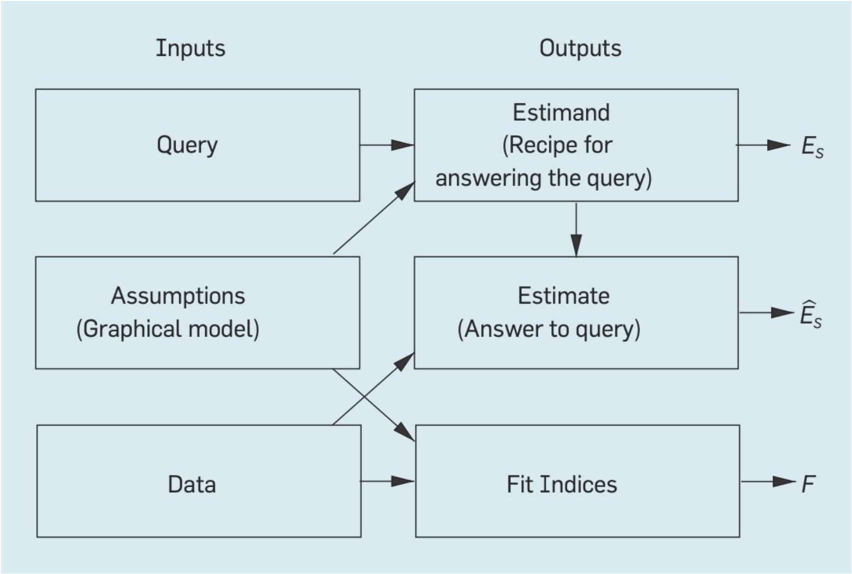
Identification: Causal effect → Observed effect conditioned on W, $\mathrm{E}[Y|T, W]$
Estimation: $\mathrm{E}[Y|T, W]$ → Propensity Score Stratification



DoWhy

Python library developed by Microsoft for causal inference based on graphs

“DoWhy is designed to highlight the critical but often neglected assumptions underlying causal inference analyses” blog microsoft



Judea Pearl : The Seven Tools of Causal Inference, with Reflections on Machine Learning

What’s next?

➔ Modelling of causal inference :

- IP weighting and marginal structural models
- the g-parametric formula, G-estimation of structural nested models
- Outcome regression and propensity scores

➔ Causal inference and machine learning, Causal discovery

➔ Causal inference from complex longitudinal data, causality in time

References

[1] J. Pearl, “The seven tools of causal inference, with reflections on machine learning,” Communications of the ACM, vol. 62, no. 3, pp. 54–60, 2019.

[2] M. A. Hernán and J. M. Robins, “Causal inference: what if,” Boca Raton: Chapman & Hill/CRC, vol. 2020, 2020.

[3] J. Pearl, “The mathematics of causal relations,” Causality and Psychopathology: Finding the Determinants of Disorders and their Cures (P. Shrout, K. Keyes and K. Ornstein, eds.). Oxford University Press, Corvallis, OR, pp. 47–65, 2010.

[4] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, and A. Zhang, “A survey on causal inference,” arXiv preprint arXiv:2002.02770, 2020.

[5] D. B. Rubin, “Causal inference using potential outcomes: Design, modeling, decisions,” Journal of the American Statistical Association, vol. 100, no. 469, pp. 322–331, 2005.

[6] “DoWhy: A Python package for causal inference.” https://github.com/microsoft/dowhy, 2019.

[7] C. Glymour, K. Zhang, and P. Spirtes, “Review of causal discovery methods based on graphical models,” Frontiers in Genetics, vol. 10, p. 524, 2019.

[8] A. Marx and J. Vreeken, “Testing conditional independence on discrete data using stochastic complexity,” arXiv preprint arXiv:1903.04829, 2019.

[9] B. Schölkopf, “Causality for machine learning,” arXiv preprint arXiv:1911.10500, 2019.

[10] J. Pearl et al., “Causal inference in statistics: An overview,” Statistics surveys, vol. 3, pp. 96– 146, 2009.

[11] P. Judea, “Causality: models, reasoning, and inference,” Cambridge University Press. ISBN 0, vol. 521, no. 77362, p. 8, 2000.

[12] J. Pearl, M. Glymour, and N. P. Jewell, Causal inference in statistics: A primer. John Wiley & Sons, 2016.

[13] J. Pearl, “The science and ethics of causal modeling,” Handbook of ethics in quantitative method- ology, pp. 383–416, 2011.

[14] H. Matute, F. Blanco, I. Yarritu, M. Díaz-Lago, M. A. Vadillo, and I. Barberia, “Illusions of causality: how they bias our everyday thinking and how they could be reduced,” Frontiers in Psychology, vol. 6, p. 888, 2015.

CentraleSupélec