Causal Inference In Statistics - An Overview, Judea Pearl

Luis Moneda

Outline

1. Paper
2. Journal
3. Author
4. Curiosities
5. Goal
6. Structure
7. Context
8. Content
9. Conclusion

Paper

Causal Inference in statistics: An overview

Published in September 2009, Statistics Surveys

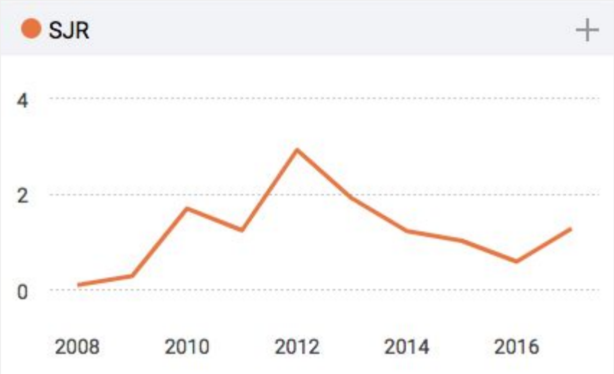
Citations: 794

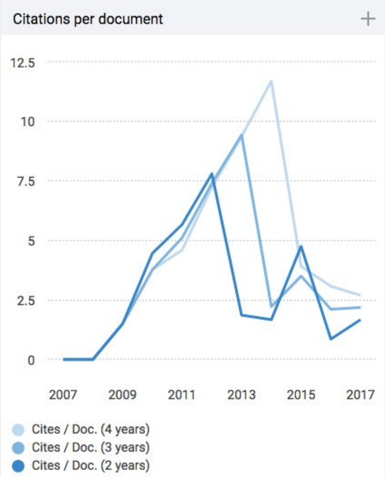
Reference: Scimago Journal & Country Rank

Journal

Statistics Surveys

It's ranked as 55th out of 217 in SJC impact factor in 2017.



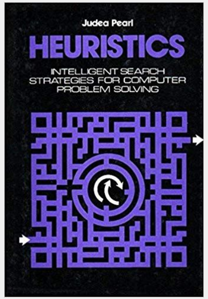


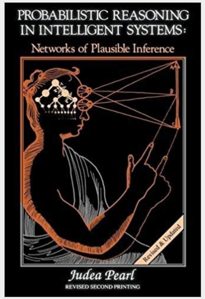
Reference: Scimago Journal & Country Rank

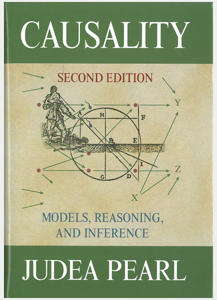
Author

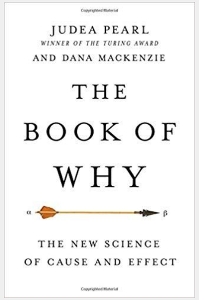
Judea Pearl

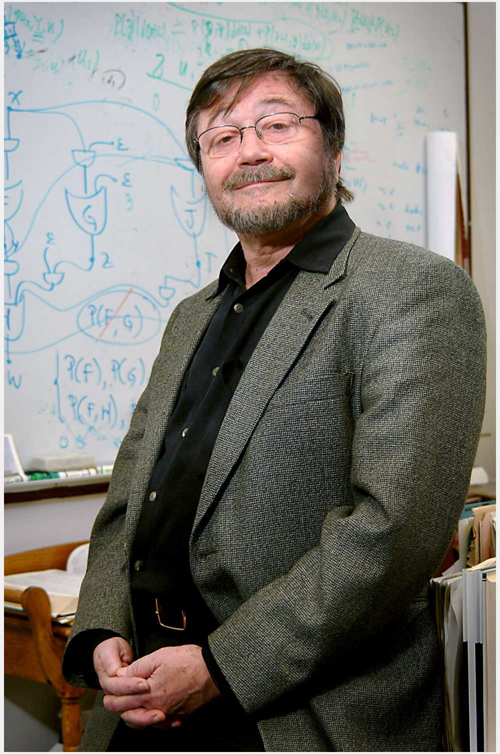
- Contributed to Search and Heuristics, Bayesian Networks and Causality;
- Turing award (2011);
- Professor and Director of the Cognitive Systems Laboratory at UCLA.











Curiosities

Curiosities:

- He warns about the fact of portions of the paper being based on his book Causality.
- From the 129 citations, 17 are from his own work.
- He has other strongly similar papers:
  - The Foundations of Causal Inference: A Review (Sociological Methodology, 2010)
  - Causal Inference in the Health Sciences: A Conceptual Introduction (Health services and outcomes research methodology, 2001)

All these papers start with a section called "From association to causal analysis" and they all seem to pursue the same goal.

Goal

Surveys the development of mathematical tools for:

1) Causal effects
2) Probabilities of counterfactuals
3) Mediation (indirect effects)

But it also contains a **comparison and analysis** about the relation between his framework and the one from potential outcomes and a way of combine them.

It compiles two decades of research on the field.

"This survey aims at making these advances more accessible to the general research community"

Structure

Very well defined, uses an index under a section called "**Contents**" just after the abstract.

- **1) Introduction**: Talks about **what** is it about, **why** the subject is relevant, presents the **structure/method,** defines the **period surveyed**, **describe the content** of each of the sections;
- **2) From association to causation:** makes it clear the differences between the two. Examples and tells how in classical statistics field the causality lacks notation.
- **3) Structural models, diagrams, causal effects and counterfactuals**: he presents his theory about structural causal models. He cites all the bases of it - previous works that contributed to it. It's a way of representing causal relations and make all the assumptions explicit. He argues it's better because it adds notation to all assumptions made and it make us able to manipulate the elements to express causal effects.

Structure

- **4) The potential outcome framework**: relate the previous section to this framework, offering a mapping between them. Offers a way to combine them and take the best of both;
- **5) Counterfactuals at work:** tools to counterfactual estimation;
- **6) Conclusions:** the same arguments from introduction and motivation.

Context

Concepts

Potential outcome

The outcome we would see under each possible treatment option (Y^n).

Counterfactual

Slightly different than potential outcomes, but often used interchangeably.

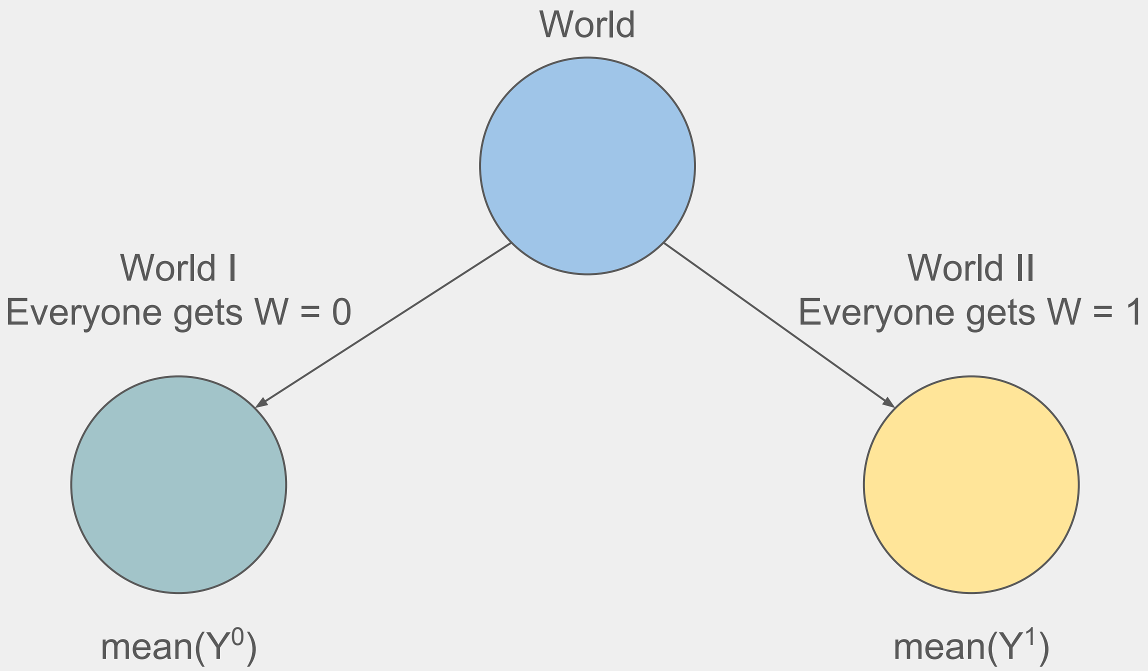
What would have happened had the action been different?

Before treatment decision is made, any outcome is a potential outcome: Y^1 or Y^0 .
After treatment there's an observed outcome Y^A and a counterfactual one Y^{1-A}.

Confounding

Anything that can impact both **treatment** and **outcome.**

Causal effect



Average Causal Effect = $E[Y^1 - Y^0]$

Causal - Randomized Controlled Trial (RCT)

It's almost like having two new worlds!

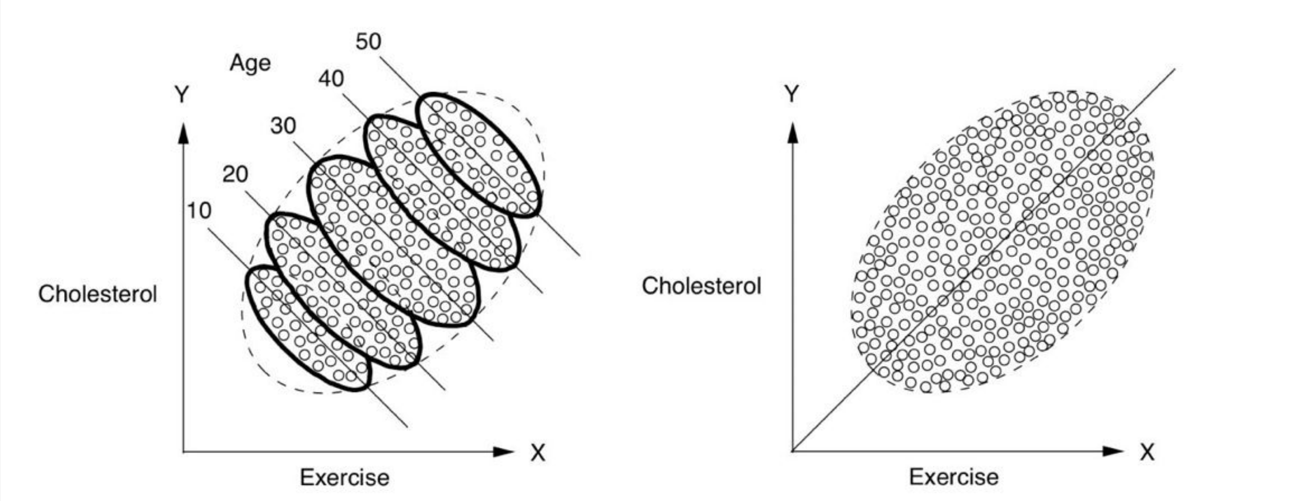
- Golden standard;
- Solves most of our problems!
- It has its own challenges, but once solved the results are robust;
- People in academia are used to do it.

Problems:

- It may not be **ethical**
- It can be **costly**

The challenge is estimating causal effect using either just **observational data** or using it with some random test data.

Motivation - Simpson Paradox



Survey Content

From association to causation - Basic distinction

Standard statistical analysis

Regression, estimation and hypothesis testing. Static description. Everything comes from the distribution alone.

There's nothing to tell what happens to a distribution if some external change occurs.

Causal Analysis

Randomization, influence, effect, confounding, intervention, explanation, attribution and so on.

It aims the dynamics of beliefs under changing conditions.

All causal statement needs assumptions untestable in observational studies!

Ramifications of the basic distinction

You can't define concepts associated with causal analysis based only using an associational definition!

Confounding is an example (Pearl, 2000a, section 6.2);

"Therefore, to the bitter disappointment of generations of epidemiologist and social science researchers, confounding bias cannot be detected or corrected by statistical methods alone;"

Standard statistics lacks notation to causal relation.

They are related, since an unambiguous notation is needed to make the causal assumptions clear.

Structural Models

A theory of causation must:

1) Represent causal question in some **mathematical language**;

2) Provide a precisa language for **communicating assumptions** under which the questions need to be answered;

3) Provide **a systematic way of answering** at least some of these questions and labeling others "unanswerable";

4) Provide a method of **determining what assumptions or new measurements** would be needed to answer the "unanswerable" questions.

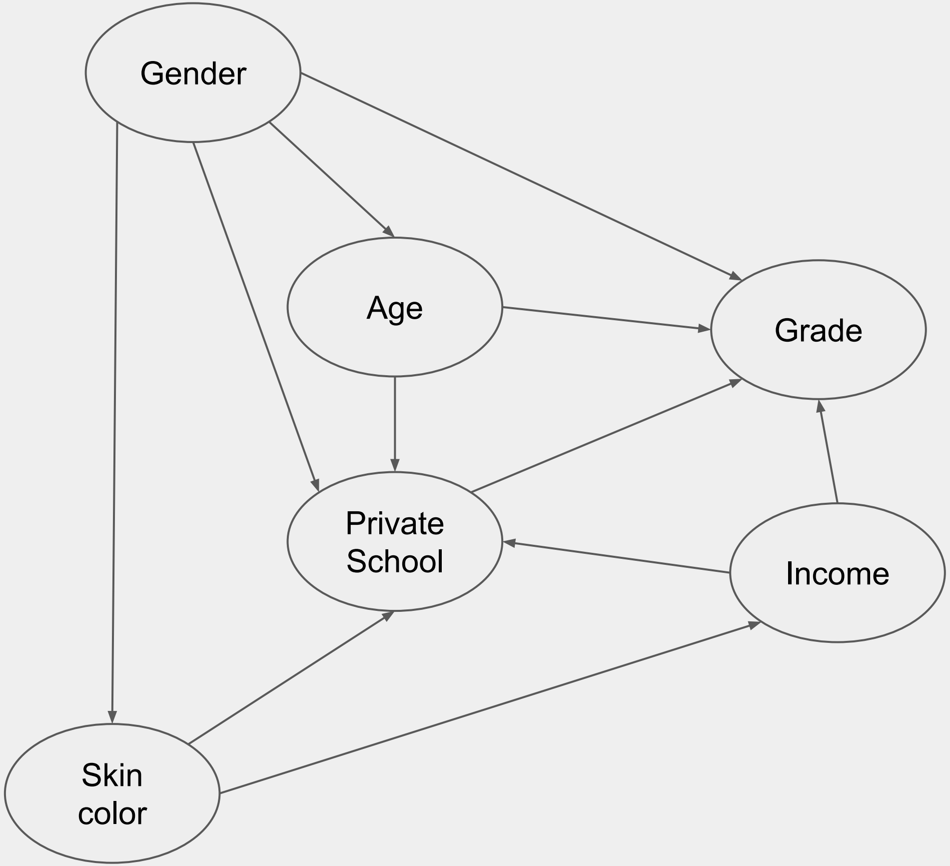
Structural Models

Combines Structural Equation models (SEM), Potential Outcomes and graphical probabilistic models. Which are described as **restricted cases** of the Structural Causal Models (SCM).

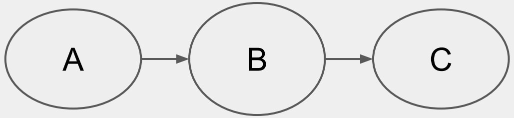
Then:

- Argues its potential is not fully utilized since some science ramifications don't use yet;
- States what should be achieve in the field;
- Introduces SEM, Graphical models, interventional notation, interventional effect estimation, how data and graph relates to estimate causal effects, the back-door criteria, relation between bayesianism and causality, an example of non-compliance in clinical trials;

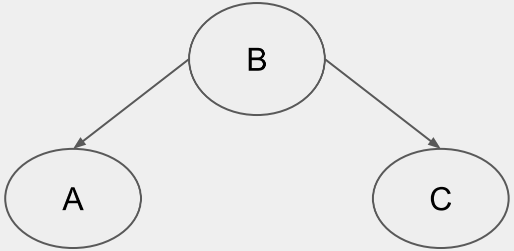
Structure Causal Model - Backdoor criteria



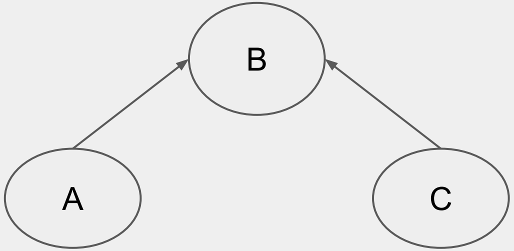
Structure Causal Model - Backdoor criteria



Chain

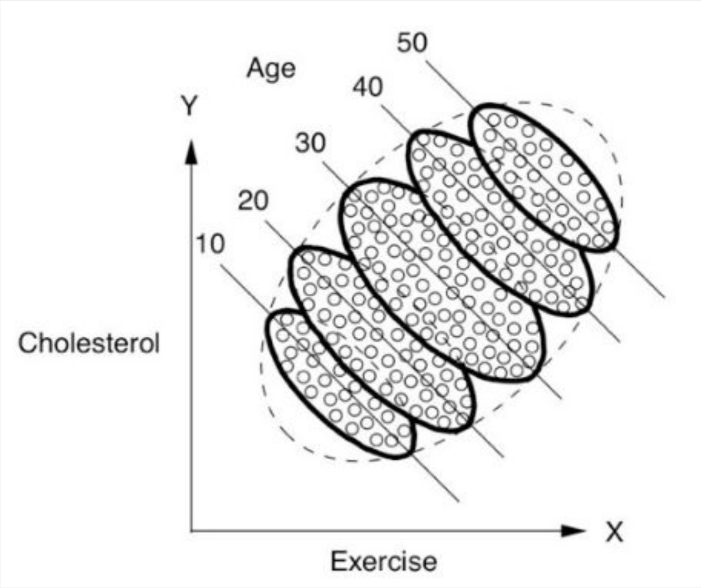


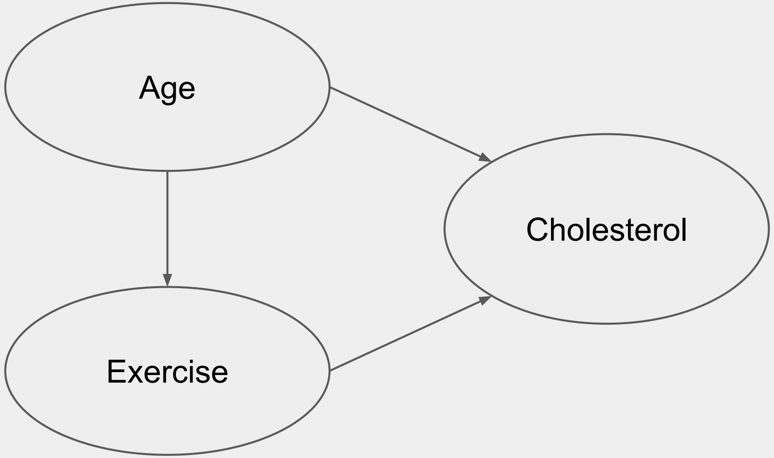
Fork



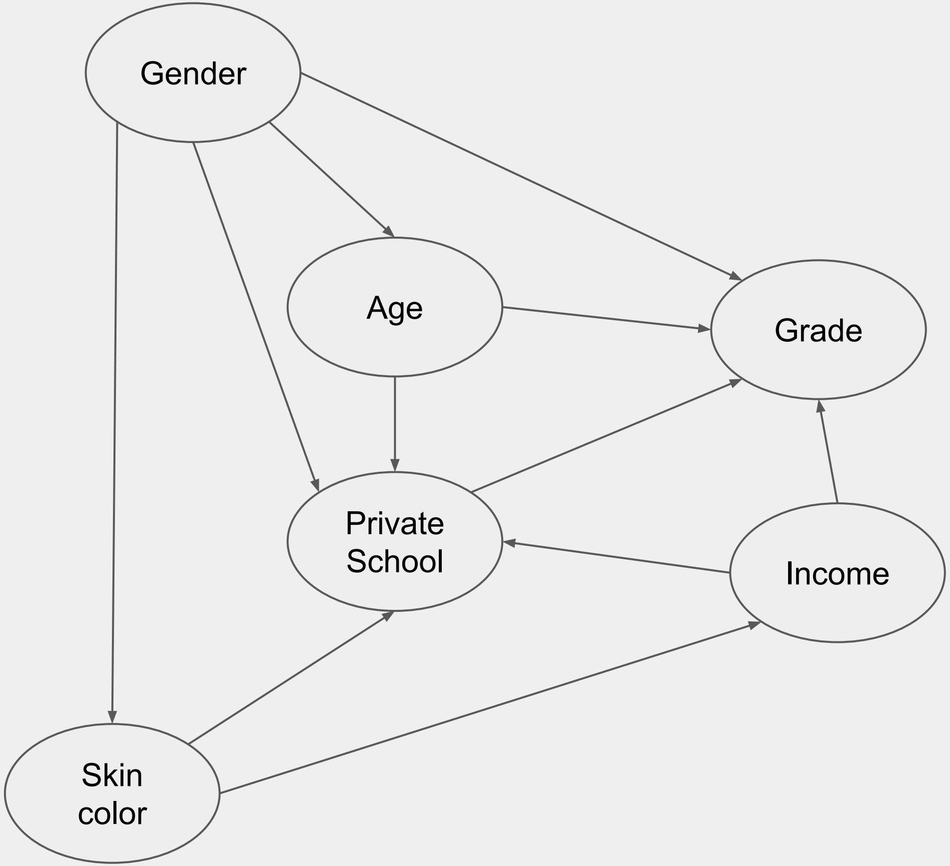
Collider

Structure Causal Model - Backdoor criteria





Structure Causal Model - Backdoor criteria



The potential outcome

- A special case;
- Lacks notation to describe all the causal relationships;
- The main drawback is the problem formulation, using PO you can't articulate the science or causal assumptions behind the problem at hand;
- Cites own works that show SCM proving the validity of potential outcome techniques;

Conclusion

Statistics is strong to **describe data and infer distribution parameters** from samples.

Causal inference has offered two main things in the last two decades: **mathematical language and machinery** to perform causal conclusions from assumptions and data.

The SCM algebraic language **overlaps with potential outcomes framework**, but it goes beyond and as a general case. When unified, **they offer a powerful and comprehensive methodology for empirical research.**

Questions?