Machine learning and causal inference (welcome)

Robert Castelo

robert.castelo@upf.edu

@robertclab

Dept. of Medicine and Life Sciences
Universitat Pompeu Fabra

upf.

Barcelona

Barcelona School of Economics
Data Science Methodology Program
Winter Term 2024-25

BSE
Barcelona School of Economics

Welcome to MLCI

What is MLCI about?

- An introduction to causal inference methods.
- Causal inference methods allow one to study the impact of some potential **treatment** (e.g., an economic intervention, a social policy, a public health measure or a laboratory experiment) on some **outcome** (e.g., gross domestic product, unemployment rate, infant mortality, tumor growth).
- The toolkit of causal inference methods is increasingly appreciated across academic fields, government, industry and non-profit organizations.

Barcelona School of Economics

Welcome to MLCI

The Telegraph News Business Sport Opinion Politics World Money Life Style Travel Culture

100-year-old woman says drink and cigarettes keep her young

A woman who toasted her 100th birthday today with a cigarette and a tot of whisky said she would also be raising a glass to 70 years as a committed smoker and drinker.

By Heidi Blake
04 March 2010 • 7:30am

Lorna Gobey celebrated her 100th birthday with a glass of whiskey and a cigarette | CREDIT: Photo: SWNS

- Should we infer that smoking and alcohol have not effect on human health?

- Counterfactual: would this woman have lived longer had she never smoked and drunk alcohol?

Barcelona School of Economics

Welcome to MLCI

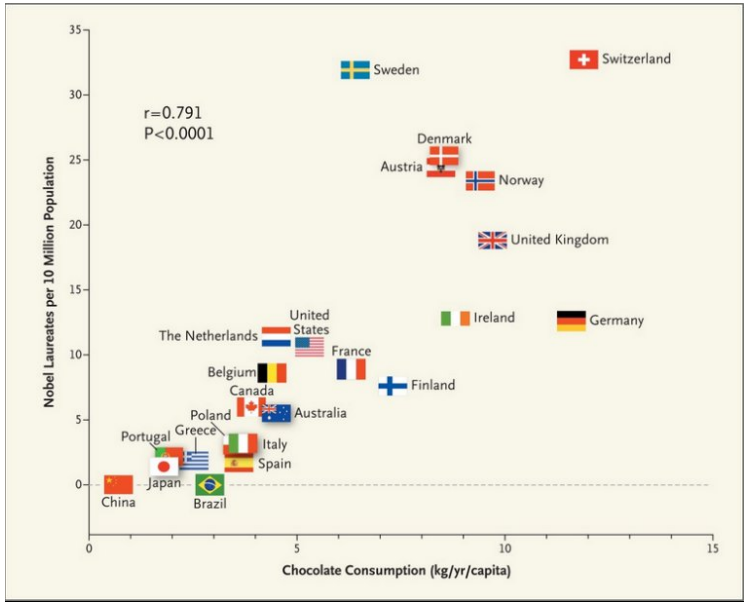
The NEW ENGLAND JOURNAL of MEDICINE

OCCASIONAL NOTES

Chocolate Consumption, Cognitive Function, and Nobel Laureates

Franz H. Messerli, M.D.

October 18, 2012
N Engl J Med 2012; 367:1562-1564
DOI: 10.1056/NEJMon1211064



"The slope of the regression line allows us to estimate that it would take about 0.4 kg of chocolate per capita per year to increase the number of Nobel laureates in a given country by 1."

Barcelona School of Economics

Welcome to MLCI

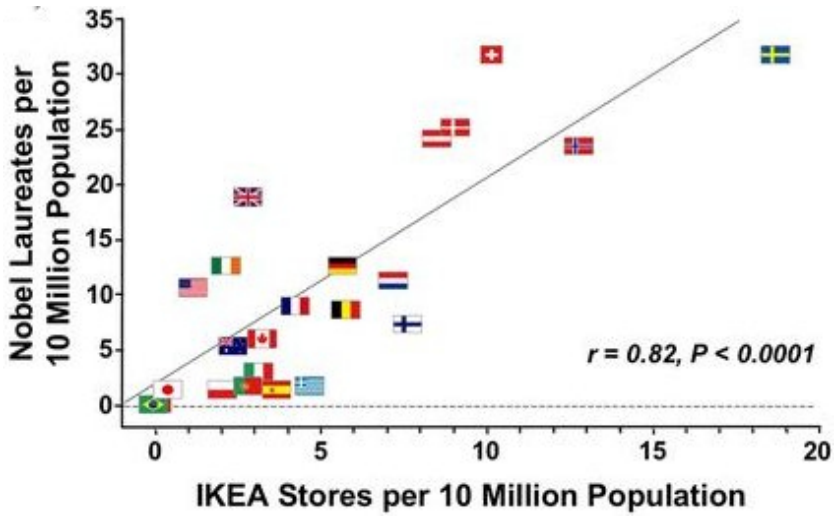
- Correlations are often spurious.
- Selection effects.
- Confounding effects.
- Causal inference is hard: many factors hamper its proper practice, and errors can lead to important misunderstandings and missteps.

Does Chocolate Consumption Really Boost Nobel Award Chances? The Peril of Over-Interpreting Correlations in Health Studies FREE

Pierre Maurage, Alexandre Heeren, Mauro Pesenti ✉ Author Notes

The Journal of Nutrition, Volume 143, Issue 6, June 2013, Pages 931–933, https://doi.org/10.3945/jn.113.174813

Published: 24 April 2013 **Article history** ▾



Welcome to MLCI

Six sessions of two hours each at 10:30am on Mondays, first five by me (Robert Castelo), plus four sessions of two hours each at 10:30am on Tuesdays by Alessandro Mascaro.

Mon Jan 13th	10:30am - 12:30pm	Randomized experiments
Mon Jan 20th	10:30am - 12:30pm	Observational studies
Mon Jan 27th	10:30am - 12:30pm	Graphical Markov models
Mon Feb 3rd	10:30am - 12:30pm	Causal inference with known structure
Mon Feb 10th	10:30am - 12:30pm	Causal inference without known structure
Mon Feb 17th	10:30am - 12:30pm	Alessandro Mascaro
Tue Feb 18th	10:30am - 12:30pm	Alessandro Mascaro
Tue Feb 25th	10:30am - 12:30pm	Alessandro Mascaro
Tue Mar 4th	10:30am - 12:30pm	Alessandro Mascaro
Tue Mar 11th	10:30am - 12:30am	Alessandro Mascaro

BSE Barcelona School of Economics

Welcome to MLCI

- The evaluation will consist of three items:
  1. Homework assignments (50%), in groups of 2 or 3 people.
  2. Final (small) project (50%), in groups of 2 or 3 people.
  3. Extra bonus (+5%), individually, find a causal claim published in the news during 2025: summarize evidence provided and propose the ideal experiment to verify it.
- Homework assignments and the final project will be submitted using GitHub. Causal claims will be submitted using a Google form.
- Office hours: online using the Zulip workspace of MLCI. Please join the Zulip workspace as soon as possible using the link at the Google Classroom.

Barcelona School of Economics

Welcome to MLCI

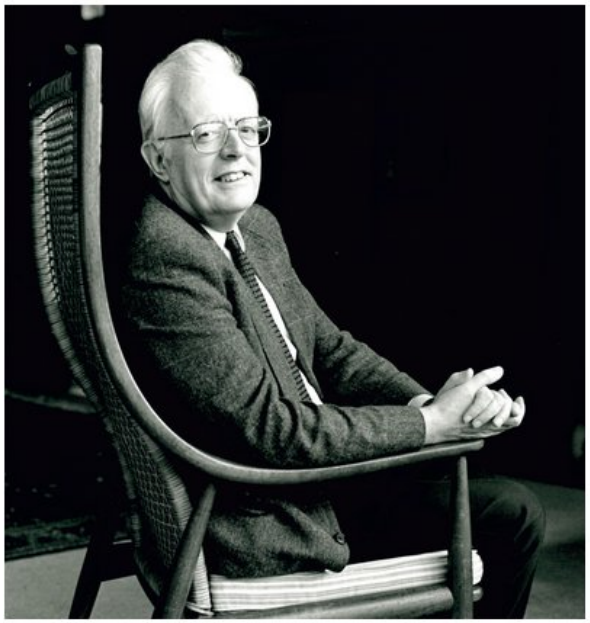
- Deadlines:
  1. Homework assignment 1: Fri Jan 24th
  2. Homework assignment 2: Fri Feb 7th
  3. Homework assignment 3: Tue Mar 4th
  4. Bonus causal question: Wed Mar 12th
  5. Final project: Tue Mar 18th

Barcelona School of Economics

In Memoriam

SIR DAVID COX, 1924-2022

We are extremely sad to share the news of the death of pioneering statistician and former Warden Sir David Cox on 18 January 2022.



It is with great sadness that we learned of the death of Sir David Cox, Honorary Fellow and former Warden, on 18 January 2022.

A pioneering statistician, David's academic legacy includes his work on binary logistic regression, the proportional hazards model which he developed in 1972, and the Cox process.

David was Warden of Nuffield College between 1988 and 1994, and an active member of the College up until his death. He will be remembered with great fondness by members of the College past and present, and his death leaves a big hole for the Nuffield community.

Obituaries for David have been published by The Wall Street Journal, The Times, *the Royal Statistical Society, St John's College, Cambridge, the Medical Research Council's Biostatistics Unit at Cambridge University, and University College London. We will provide information about a memorial service in due course.*

Photograph (c) Jane Bown.

https://www.nuffield.ox.ac.uk/news-events/news/sir-david-cox-1924-2022

In the following interview made a few months before David Cox died at the age of 97, he talks (minute 18:58) about the importance in statistics of formulating good questions and his perception of the use of the term *causality*.

Barcelona School of Economics