Causal Inference 2.0

Wassim Tarraf, PhD

MCUAAAR ANC

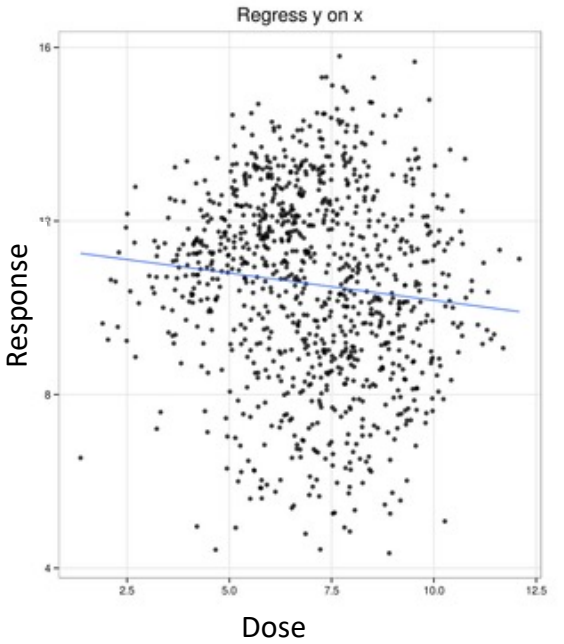
01/19/2021

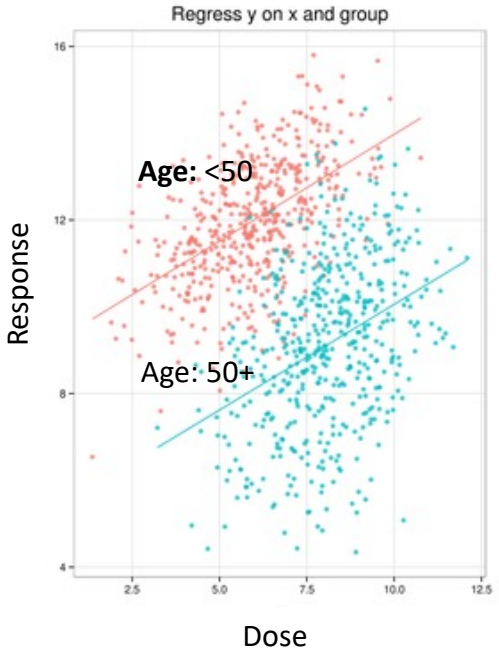
Plan

- A very brief introduction to causation
- Modeling strategies
- Focusing on matching
- How to Matching
- Further readings

Correlation does not imply causation

- The rooster and the sun (crows before sunrise), breathing and death (every breathing being dies)
- Variables:
  - Endogenous (effects): determined by the model -> “variables that show differences we wish to explain”
  - Exogenous (causes): determined by factors outside the model -> “variables used to explain the differences”
- For two variables X and Y
  - X can cause Y
  - Y can cause X (reverse directional) – playing violent video games increases tendency for violence vs. individuals with more tendency for violence play violent video games
  - Both Y and X are caused by another variable C (common cause) – Sleeping with shoes on causes a headache!
  - X causes Y and Y causes X (bidirectional)
  - X and Y covary coincidentally
- Potential for causal relationship opposite to or in the absence of observable correlation (e.g. confounding, threshold effects, noise oversaturation, non-linear).





Petersen, T. (2001). Endogeneity: Methodology. International Encyclopedia of the Social & Behavioral Sciences. N. J. Smelser and P. B. Baltes. Oxford, Pergamon: 4511-4513.

Simpson's Paradox

Data as a whole trends (between two variables) in one direction; but different patterns emerge after subgrouping on a third variable

UC Berkeley gender bias

	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
Total	12,763	41%	8,442	**44%**	4,321	35%

Department	All		Men		Women	
	Applicants	Admitted	Applicants	Admitted	Applicants	Admitted
A	933	64%	***825***	62%	108	**82%**
B	585	63%	***560***	63%	25	**68%**
C	918	35%	325	**37%**	***593***	34%
D	792	34%	417	33%	375	**35%**
E	584	25%	191	**28%**	***393***	24%
F	714	6%	373	6%	341	**7%**
Total	4526	39%	2691	45%	1835	30%

Bickel, Peter J., Eugene A. Hammel, and J. William O'Connell. "Sex bias in graduate admissions: Data from Berkeley." *Science* 187.4175 (1975): 398-404.

What is causal inference?

- Methods used to establish cause and effect
- Impact of (“events”,“choices”, ”programs”, “policies”, etc…) exposures on outcomes
- Examples:
  - Does head start reduce incarceration?
  - Does diabetes increase the incidence of ADRD
  - Does Medicaid eligibility expansion increase use of preventive services?
  - Do health warnings and labeling on cigarettes reduce smoking prevalence?
- Includes a range of modeling techniques

Fundamental problem for causal inference

Potential outcome

- Treatment T
- Outcome Y
- Individual i

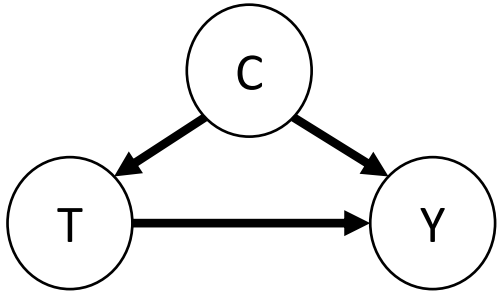
Individual Treatment Effect:
Yi(T)-Yi(nT)

Average Treatment Effect: $E[Yi(T)-Yi(nT)] = E[Y(T)]-E[Y(nT)]$

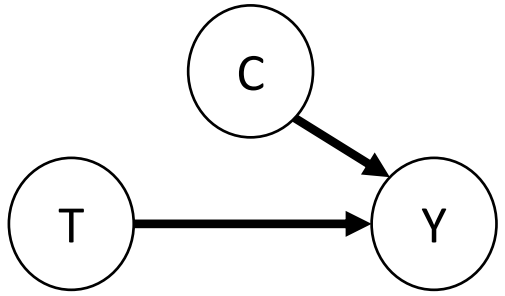
Individual Level | Population Level

ATE Breaks in the presence of confounding:

$E[Y(T)]-E[Y(nT)] \neq E[Y|T]-E[Y|nT]$



Randomization



- Randomness vital for causal effects

Re-establish ATE when confounding:

$E[Y(T)]-E[Y(nT)] = E[Y|T]-E[Y|nT]$

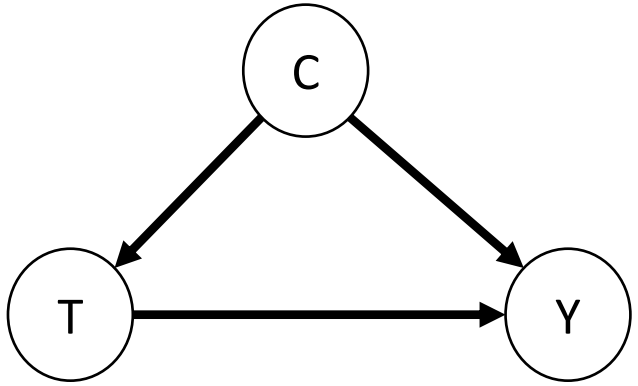
- Experimental data (Key benefit is treated and control groups are randomly different on observed and unobserved characteristics)
  - Hard in the social sciences: Cost, feasibility, ethics, etc…
  - Possible: the Oregon Medicaid Experiment or the RAND health insurance experiment
- Non-Experimental data (observational)
  - Largely the norm in the social sciences
  - Major issue with endogeneity -> **CHOICE** by people making decisions about what is best
  - For a causal relationship to exist the decisions about the exposures have to be made ”independent of the potential outcome” of interest
  - you have to rely on “observed covariates”

SOLUTION

ADJUST FOR CONFOUNDERS

The ”standard” modeling strategy

- Interest in linking T and y
- Random sampling from a population of interest
- Probe how y varies with changes in T
- Questions:
  - What if y is affected by factors other than T; how should we handle that?
  - What is the functional form connecting the two variables?
- Most of you/us stop there and devise a modeling strategy. However:
  - How can we attribute the variation between y and T to something other than correlation
  - What if T is confounded by other factors; how should we handle that??



Causal Inference Techniques

- Potential outcome model
- Graphical Models: A graphical representation of a causal chain (nodes and arrows)
- Difference in Difference, Regression discontinuity, and IV models
- Panel data
- Synthetic control

- Our focus today is on: **Matching techniques for causal inference**

See: Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Matching

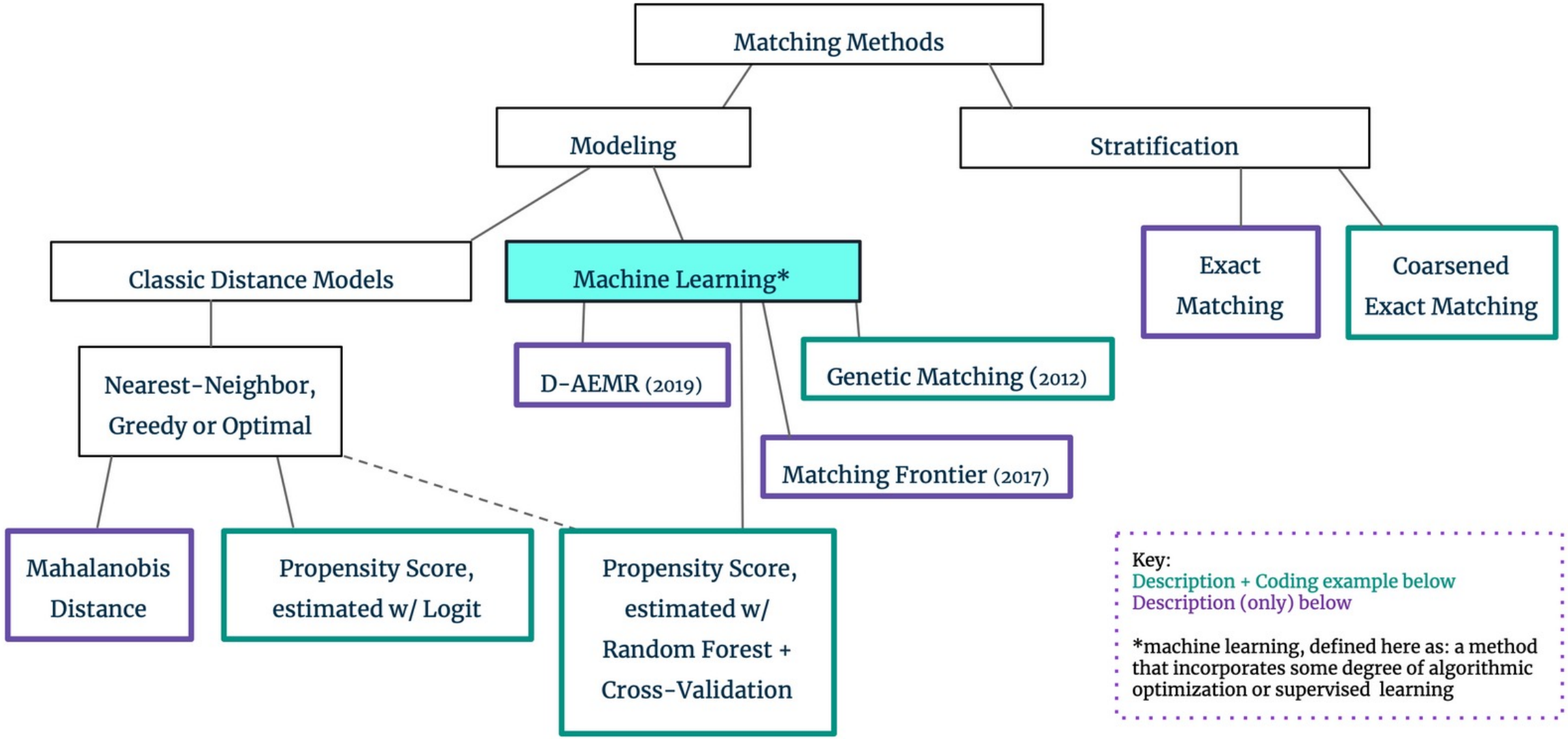
- Randomized Experiments: Key benefit of it is treated and control groups are randomly different on observed and unobserved characteristics
- Matching methods: focus on how to replicate the above conditions using observed covariates in the absence of randomized experiments.
- Matching: Approximating distribution of covariates in treated and untreated groups
  - Complementary of regression techniques
  - Overcome weaknesses of regression and selection techniques in the absence of sufficient overlap
  - Straightforward diagnostics to assess performance

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Estimators

- Standard Regression estimators
  - Sensitive to dissimilarities in covariates distribution in the treated and control
  - Nonparametric estimators -> but presents some problems as well
- Matching estimators
  - Direct adjustment for all covariates; complex as # of covariates increases
- Propensity score (PS) estimators
  - “Adjustment” on a single index
    - Weighting on the inverse of the PS: balance the sample of treatment and controls by up/down weighting sample observations.
    - Stratification on the PS: (1) Divide the sample in M stratas with approximate probability of treatment (2) Estimate Average Treatment Effect within each strata; (3) Calculate the overall ATE/ATT as the average of of averages
    - Regression on the PS: Expected value of outcome given T conditional on the PS
    - Matching on the PS: see below

Matching



See: https://humboldt-wi.github.io/blog/research/applied_predictive_modeling_19/matching_methods/

Quantities of interest

- ATE: Average Treatment Effect on all individuals (in treatment and control groups): $\frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0))$
- ATT: Average Treatment among the Treated (effect for those in the treated group): $\frac{1}{N_t}\sum_{i:W_i=1}^{N}(Y_i(1) - Y_i(0))$
- Other:
  - Conditional ATE
  - Conditional ATT
  - FSATT (ATT in feasible samples; in the absence of good control matches)

See: https://humboldt-wi.github.io/blog/research/applied_predictive_modeling_19/matching_methods/; https://docs.zeligproject.org/index.html

Assumptions

- Stable unit treatment value assumption (SUTVA)
  - treatment same for all *i*s
- Unconfoundedness/ignorability/”no hidden bias):
  - Treatment assignment is independent of outcome conditional on covariates
  - Not directly testable (falsification)
    - Estimate effect of a treatment with known null effect; e.g. split the control group and estimate within control group for comparison
    - Estimate effect on an outcome with known null
  - Worry about unobserved covariables that are not related to observed covariates.
- Overlap:
  - Matching is only valid in areas of common support
  - In case of lack of overlap
    - Recognize potential for imprecision
    - Attempt to correct for overlap
      - Discard bad matches
      - Focus on matches with acceptable difference in PS

Process

- Defining a distance measure -> to estimate closeness
- Doing the matching -> based on the measure of closeness defined in step 1
- Diagnosing the matching -> Assess quality of matched sample
- Estimating the treatment effect after matching.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Process

- Defining a distance measure -> to estimate closeness
  - How to choose variables to satisfy ignorability condition (no unobserved differences between treatment and control conditioned on observed covariates)
  - Include variables related to both treatment assignment and outcome
    - No cost for including variables that are not associated with treatment
    - Potential increase in variance for including variables not associated with outcome
    - In large samples ~50-100; prohibitive in small samples
    - Include small”er” set of variables known to be associated with outcome – check balance on a larger set of covariates
    - Use of substantive knowledge and previous literature critical
    - Do not include variables that may have been affected by the treatment
    - Be careful regarding variables that are highly predictive of treatment

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Process

- Defining a distance measure -> to estimate closeness
  - Propensity Score
  - Linearization of propensity score
  - Other (omitted here; don't work as well with high dimensional X, problems in the presence of non-normally distributed variables, etc…)

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Propensity Score Matching

- Propensity scores are predicted values generated from the estimation of a model for the conditional probability of treatment
- Bounded prediction (0-1) that collapses a set of covariates (to match on) into a single index.
- Comparisons between the treatment and the control groups are based on distance using this index.
- For binary treatment estimation standard is logistic regression (other non-parametric models can be used e.g. random forests; lots of simulations to see what works best under what conditions).
- Focus on post estimation balance of the predicted scores.
  - Fit criteria not helpful
  - Standard issue of multicollinearity not a concern
  - Variable selection methods not useful
  - Imbalance should examine original variables and functions of those variables; in case of imbalance inclusion of these is recommended
  - Limited evidence that treatment effects robust to potential mis-estimation of the PS

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Process

- Doing the matching -> based on the measure of closeness defined in step 1
  - Nearest neighbor matching
    - Optimal matching
    - Ratio matching
    - With and without replacement
  - Stratification
  - Weighting
    - Inverse probability of treatment
- Common support: overlap in the distribution
  - Lack of common support has consequences for reliability in estimation depending on the Estimand

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Process

- Diagnosing the matching -> Assess quality of matched sample
  - Assess the balance of lower dimensional summaries of covariables
  - Balance assessment should reflect the method (matching vs. stratification vs. weighting)
  - Numerical and graphical diagnostics
    - SMD of the score
    - Ratio of variance of the PS in treated and control groups
    - SMD of the covariates
    - P-values not valid

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Process

- Estimating the treatment effect after matching
  - Outcome analyses on the matched, stratified, weighted samples
  - Accounting for matched pairs not necessary
  - Conditioning on variables used in PS estimation sufficient
  - Weights should be included when matching with replacement or when ratio matching (i.e. not 1:1 is used)
- Variance estimation:
  - Standard methods for inference (robust, cluster-robust,…)
  - Debate: to bootstrap or not to boostrap
  - Different procedures have been recommended for specific matching scenarios

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1.

Resources

Resources

Review specific to matching

Matching methods for causal inference: A review and a look forward

Survey books on causal inference

Counterfactuals and Causal Inference: Methods and Principles for Social Research

Open science books on causal inference

Causal Inference: the Mixtape

Applied Causal Analysis (with R)

Books on matching

With R code

Practical Propensity Score Methods Using R

With Stata code

Propensity Score Analysis Statistical Methods and Applications