EMORY | COLLEGE OF ARTS AND SCIENCES

Estimation

Intro. to causal inference | SPSP 2023 Annual Convention

Wen Wei Loh (wen.wei.loh@emory.edu)
Department of Quantitative Theory and Methods (QTM)

February 25, 2023

Causal assumptions before estimation

Outcome regression adjustment

What covariates not to adjust for

Propensity score estimation

Regression adjustment for the propensity score

Inverse weighting using the propensity score

Trying them out in R

Summary

Causal assumptions before estimation

Can these causal assumptions be justified?

- Reverse causation ($X \leftarrow Y$) can be ruled out.
- Treatment: conceptually and practically manipulable.
- Causal consistency and SUTVA hold.
- Well-defined average treatment effect (ATE) estimand.
- Precisely measured baseline covariates L sufficient to adjust for all confounding
- I.e., conditional exchangeability or no unmeasured confounding to hold given L.

We can now proceed with estimating the ATE!

Outcome regression adjustment

Outcome regression adjustment

- Recall that under conditional exchangeability given L:

$$\mathrm{E}[Y(1) - Y(0)|L] = \mathrm{E}[Y|X = 1, L] - \mathrm{E}[Y|X = 0, L].$$

- *Left*: conditional average difference of potential outcomes.
- Fundamental problem of causal inference.
- *Right*: model and estimate predicted outcomes using the observed data.

Outcome regression adjustment

- When Y is continuous, we can use linear regression; e.g.,

$$\mathrm{E}[Y|X, L] = a + bX + cL. \tag{1}$$

- This is an example of an outcome regression model.
- Estimating a *main effect* for treatment X (coefficient b).
- *Statistically control* for baseline covariates L.
- How to interpret the coefficient of X?
- Can we use b to estimate the ATE?

Linear and additive treatment effect

- The coefficient of X (b) is consistent for the ATE if:
  1. no treatment-covariate interaction terms (i.e., no effect modification or moderation) are included; and
  2. the relationships between Y and covariates L are correctly specified.
- Then we can use routine ordinary least squares or maximum likelihood estimators of b to estimate the ATE.

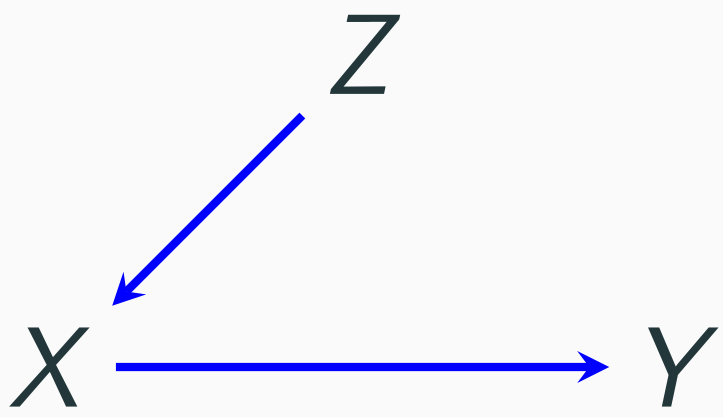
What covariates not to adjust for

Adjusting for non-confounders

- **Recap**. when estimating the ATE of X on Y, we should:
  - adjust for confounders on backdoor paths between X and Y;
  - not adjust for colliders along non-causal paths between X and Y; and
  - not adjust for variables (or their descendants) affected by X.

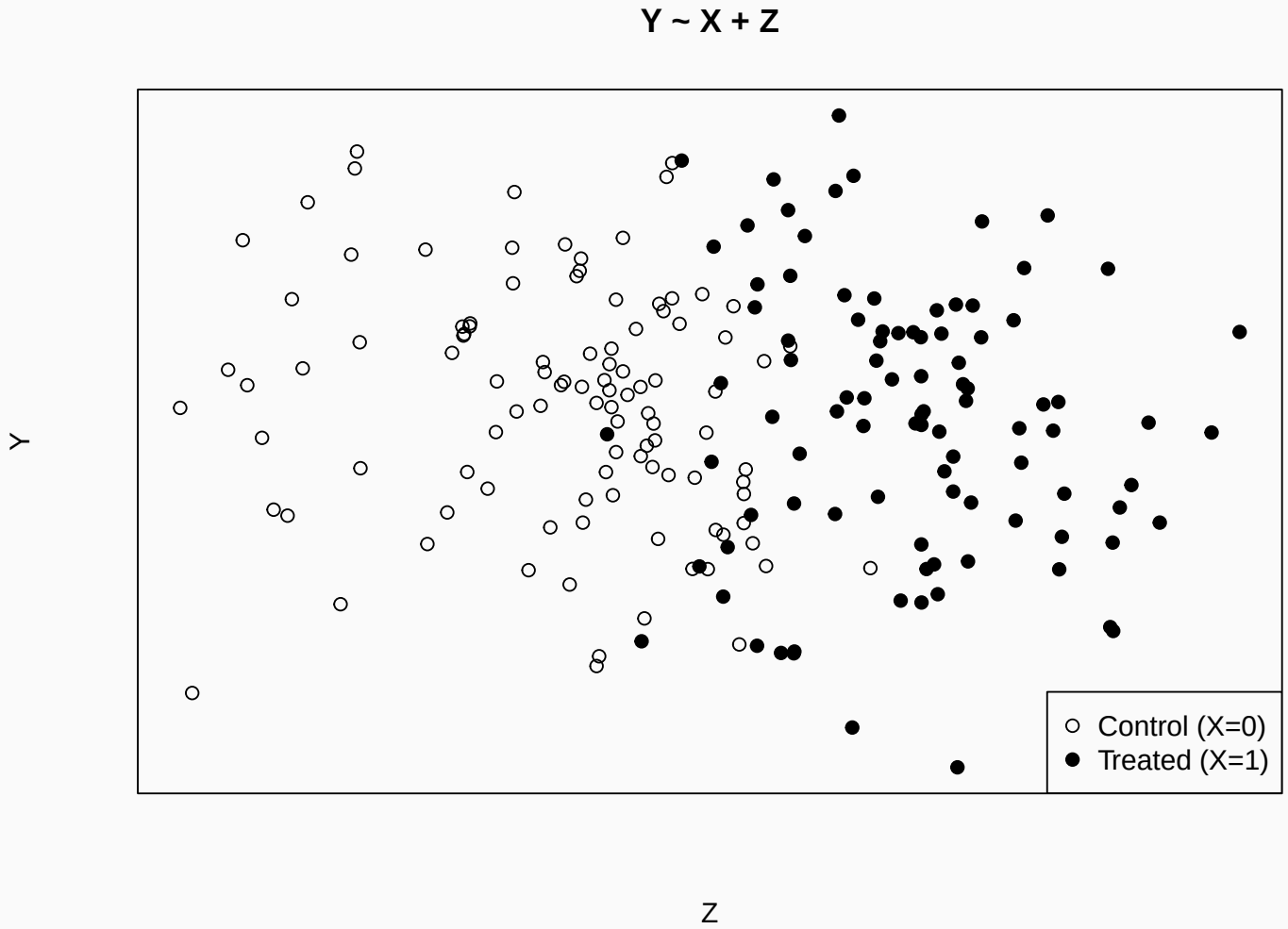
Adjusting for non-confounders

- Suppose Z predicts treatment X but not outcome Y.
- E.g., proximity of your hotel to a coffee shop (Z) may affect whether you drink a coffee (X) but is unlikely to affect your tiredness (Y).



- Do we need to adjust for Z?

Adjusting for non-confounders



Adjusting for non-confounders

- We are using the regression model

$$\mathrm{E}[Y|X, L] = a + bX + cZ.$$

  to compare outcomes Y with different X, but same Z.
- When the treatment groups have little (or no) overlap in the covariates:
  - Multicollinearity: strongly correlated predictors of outcome;
  - Non-positivity: positivity assumption may be violated;
  - Extrapolation: predict counterfactual outcome under unobserved values of Z.

Adjusting for non-confounders

- Including treatment-only predictors can lead to unstable estimates.
- Can arise even with models that fit the data well [Vansteelandt and Daniel, 2014].
- Adverse impact can be difficult to detect, especially with many baseline covariates.
- Treatment-only predictors are unnecessary for confounding adjustment and should be avoided [VanderWeele, 2019].

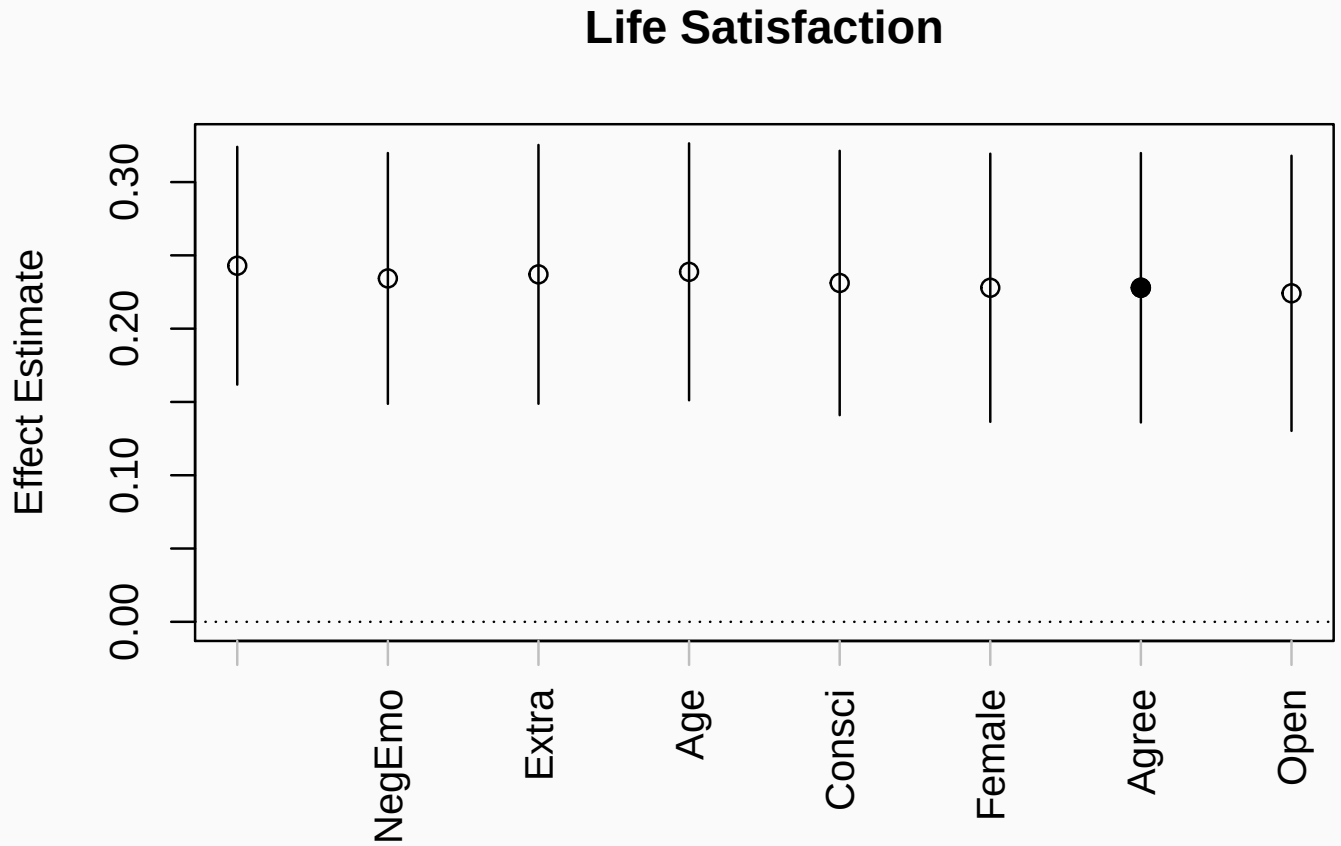
Data-driven covariate selection

- Suppose a set of covariates L has been deemed sufficient for conditional exchangeability to hold.
- Among these may be treatment-only predictors *unnecessary for confounding adjustment*.
- Interest may be in assessing how the effect estimate changes as different covariates are adjusted for.
- One approach: stability-based evaluation [Loh and Ren, 2023].

Data-driven covariate selection

1. Order the covariates in decreasing priority for confounding adjustment.
    - E.g., (i) Negative Emotion; (ii) Extraversion; (iii) Age; ...
2. Calculate the effect estimator using each nested subset.
    - E.g., (i) {Negative Emotion};
      (ii) {Negative Emotion, Extraversion};
      (iii) {Negative Emotion, Extraversion, Age}; ...
3. Select the subset that yields the most stable effect estimator.

Data-driven covariate selection



Propensity score estimation

Propensity score

- Outcome regression models are prone to misspecifying outcome-covariate relations and extrapolation.
- Can be avoided by modeling **treatment-covariate** associations instead.

Propensity score

- Another approach to account for non-random treatment selection: propensity score [Rosenbaum and Rubin, 1983].

$$e(L) = p(L) = \Pr(X = 1|L).$$

- It is the probability (or "propensity") of selecting treatment given baseline covariates.
- Values can be between 0 and 1.
- Check: In a randomized experiment where assignment to treatment or control is equally likely, $e(L) =?$

Propensity score

- The propensity score summarizes all covariates in L.
- E.g., suppose a treated and an untreated individual both have the same value of the propensity score; e.g., $e(L) = 0.4$.
- Then they are equally likely to have selected treatment – as if they were pseudo-randomized.
- They need not have the same values of multiple L – just the same value of $e(L)$.

Propensity score

- Under a correctly-specified model for $e(L)$ and assuming positivity holds, individuals with the same value of $e(L)$ are conditionally exchangeable.
- Permits "like with like" comparisons between treated and untreated individuals.
- The covariate distributions between treated and untreated individuals are balanced given the propensity score [Austin, 2011, West et al., 2014].

Propensity score estimation

- In practice, propensity scores are typically unknown in observational studies.
- They can be estimated using e.g., *logistic regression*.
- E.g.,

$$e(L) = \text{logit}\{\Pr(X = 1|L)\} = a + bL. \tag{2}$$

- Statistics: $\text{logit}\{p\} = \log\left\{\frac{p}{1-p}\right\}$.
- How to use the propensity scores to estimate the ATE?

Regression adjustment for the propensity score

Regression adjustment for the propensity score using `lm`

- Add the propensity score (PS) as a covariate when regressing Y on X. E.g.,

$$\mathrm{E}[Y|X, e(L)] = a + bX + c\, e(L).$$

- E.g., in the `lm` function in `R`:

```
lm(Y ~ X + PS, ...)
```

Adjusting for the propensity score

- No need to model the outcome-covariate associations; instead requires correctly modeling the propensity score.
- Not prone to extrapolation bias or covariate non-overlap [Vansteelandt and Daniel, 2014]:
- Standard errors and confidence intervals using non-parametric bootstrap.

Inverse weighting using the propensity score

Inverse probability weighting

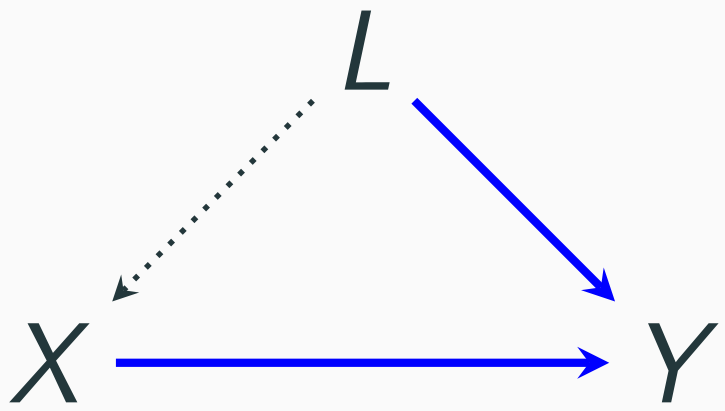
- Construct weights that eliminate confounding due to L.
- For each individual, define their inverse probability weight (IPW) as:

$$W = \begin{cases} \frac{1}{e(L)}, & X = 1; \\ \frac{1}{1 - e(L)}, & X = 0. \end{cases} \tag{3}$$

- W is the inverse of an individual's probability to receive their *observed treatment* given L.
- Aside: also termed *inverse propensity score weight* or *inverse probability of treatment weight* (IPTW).

Intuition behind IPW

- IPW creates a pseudo-population in which treatment and control groups have the same distribution on covariates L.
- The links between treatment X and covariates L are broken in such a pseudo-population.
- This eliminates confounding due to L.



IPW estimation using `lm`

- Include W as weights when regressing Y on X. E.g.,

$$\mathrm{E}[Y|X, e(L)] = a + bX.$$

- E.g., in the `lm` function in `R`, use the `weights` argument:

```
lm(Y ~ X, weights = W, ...)
```

Trying them out in R

Activity!

Try analyzing a simulated dataset in R.

https://github.com/wwloh/spsp2023-causal

Dataset name: spsp2023-causal-lab-Data1.csv

- Outcome: Y
- Treatment: TRT
- Covariates: L1 and L2
- True causal effect = 0

Activity!

Using propensity scores.

1. Regress treatment (TRT) on covariates (L1 and L2) using logistic regression.
2. Calculate predicted propensity scores (PS).
3. Regress outcome Y on treatment (TRT) and PS.
4. Calculate inverse probability weights W using PS.
5. Regress outcome Y on treatment (TRT) with weights W.

https://github.com/wwloh/spsp2023-causal

Activity!

True effect = 0	No covariates	Outcome Reg.	Reg. adjustment for PS	IPW
Estimate	1.10	0.00	0.00	-0.09
SE	0.03	0.02	0.02	0.09
95% CI	(1.04, 1.16)	(-0.03, 0.04)	(-0.04, 0.03)	(-0.26, 0.09)

Summary

Comments on IPW estimator

- Neither sufficient nor necessary to predict X well – focus is on covariate balance.
- Check covariate imbalance [Kainz et al., 2017].
- SEs and CIs using non-parametric bootstrap; or a robust (sandwich) variance estimator.
- Individuals with $X \approx e(L)$ can have disproportionately large estimated weights W. This can lead to unstable estimates.
- IPW estimators tend to have higher variability relative to regression adjustment.

Outcome regression vs. propensity scores

Outcome regression	**Propensity scores**
Requires correctly modeling outcome Y vs. treatment X and covariates L.	Requires correctly modeling treatment X vs. covariates L.
No propensity score (or "selection") model is needed.	No outcome model is needed.

Summary

- Consider different estimation methods: regression adjustment or propensity scores?
- Revisit your assumed estimation models: include non-linear terms, e.g., covariate-covariate interactions?
- Avoid treatment-only predictors for confounding adjustment!
- Consider data-driven covariate selection methods; e.g., stability-based evaluation [Loh and Ren, 2023].

Thank you! Any comments or questions: wen.wei.loh@emory.edu

References

P. C. Austin. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3):399–424, 2011. doi: 10.1080/00273171.2011.568786.

K. Kainz, N. Greifer, A. Givens, K. Swietek, B. M. Lombardi, S. Zietz, and J. L. Kohn. Improving causal inference: Recommendations for covariate selection and balance in propensity score methods. *Journal of the Society for Social Work and Research*, 8(2):279–303, 2017. doi: 10.1086/691464. URL https://doi.org/10.1086/691464.

W. W. Loh and D. Ren. Data-driven covariate selection for confounding adjustment by focusing on the stability of the effect estimator. *Psychological Methods*, In print, 2023. doi: 10.31234/osf.io/zkdqa. URL psyarxiv.com/zkdqa.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

T. J. VanderWeele. Principles of confounder selection. *European journal of epidemiology*, 34(3):211–219, 2019. doi: 10.1007/s10654-019-00494-6.

S. Vansteelandt and R. M. Daniel. On regression adjustment for the propensity score. *Statistics in Medicine*, 33 (23):4053–4072, 2014. doi: 10.1002/sim.6207.

S. G. West, H. Cham, F. Thoemmes, B. Renneberg, J. Schulze, and M. Weiler. Propensity scores as a basis for equating groups: Basic principles and application in clinical treatment outcome research. *Journal of Consulting and Clinical Psychology*, 82(5):906–919, 2014. doi: 10.1037/a0036387.