STATISTICAL HORIZONS

Matching and Weighting for Causal Inference with R

Stephen Vaisey, Ph.D.

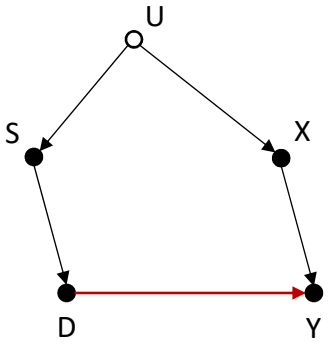
Upcoming Seminar:
July 27-30, 2021, Remote Seminar

Matching and Weighting for Causal Inference Using R

Stephen Vaisey

1

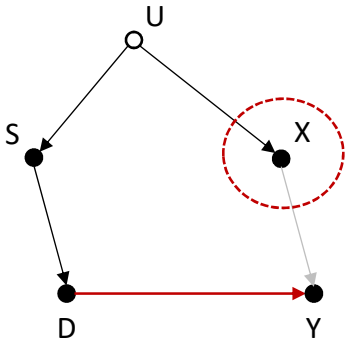
DAGs: Regression and matching



Backdoor path: is there another way to connect D and Y that doesn't pass through a "collider" (a point with two arrows pointing in)? If so, then you have to break the path to identify the effect of D on Y.

2

Regression
(identification by adjustment)

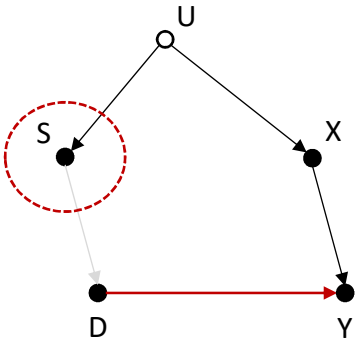


Regression intervenes here, breaking the backdoor path by removing the effect of X on Y.

3

3

Matching/weighting
(identification by balancing)



Matching and weighting break the backdoor path here, by balancing on S so there is no longer any difference in S between D=0 and D=1.

4

4

Map of the rest of the course

1. Theoretical background
2. Exact matching
3. Propensity score methods (parametric and semi-parametric)
4. Non-parametric methods
5. Parametric regression with preprocessed data
6. Extensions

5

5

The logic of propensity scores

Because exact matching is impossible when *S* comprises many variables, propensity scores allow us to summarize *S* in a single, continuous variable. This allows comparing "apples to apples" as long as we are comfortable with "appleness" as defined by the propensity score.

6

6

Two big questions

How do I estimate propensity scores?

- logistic regression
- generalized boosted modeling
- covariate balancing p-score
- any other classifier

choice of metric

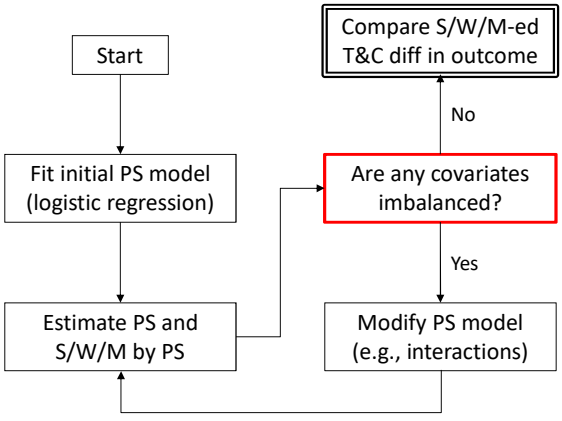
- log-odds
- probability

What do I do with them?

- stratification
- weighting
- matching

7

Traditional workflow



Adapted from Diamond and Sekhon 2014
S/W/M = stratify, weight, or match

8

Estimating p-score by logistic regression

```
psmod <- glm( mbsmoke ~ mmarried + alcohol + mrace + fbaby + mage + I(mage^2) + medu + nprenatal ,
              data = d ,
              family = binomial )
summary( psmod )
```

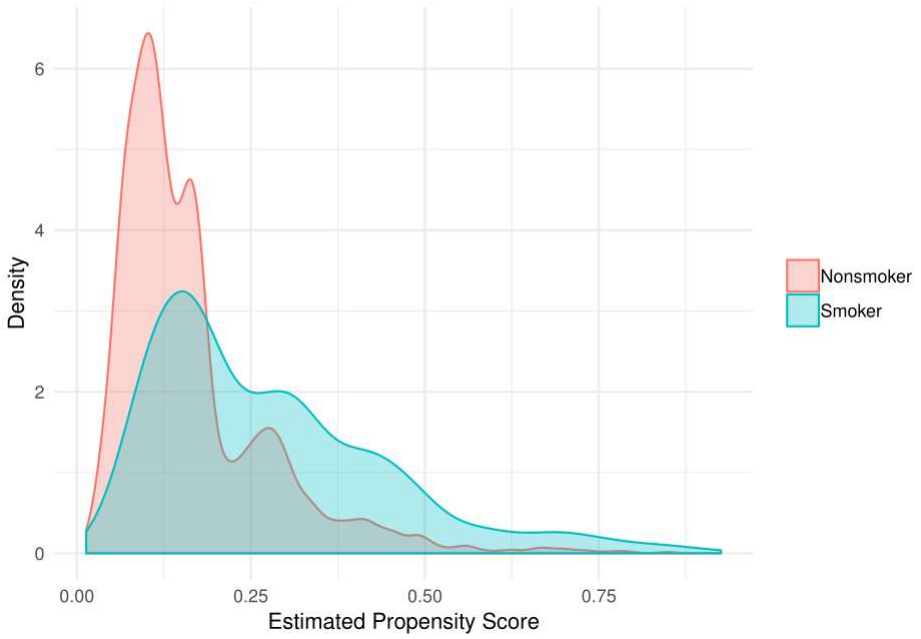
```
Call:
glm(formula = mbsmoke ~ mmarried + alcohol + mrace + fbaby +
    mage + I(mage^2) + medu + nprenatal, family = binomial, data = d)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9437  -0.6185  -0.4869  -0.3574   2.5515

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.043472   0.823455  -3.696 0.000219 ***
mmarried1   -1.235271   0.099901 -12.365  < 2e-16 ***
alcohol1     1.566114   0.184948   8.468  < 2e-16 ***
mrace1       0.666586   0.118360   5.632 1.78e-08 ***
fbaby1      -0.405226   0.090661  -4.470 7.83e-06 ***
mage         0.311710   0.064479   4.834 1.34e-06 ***
I(mage^2)   -0.005909   0.001193  -4.952 7.35e-07 ***
medu        -0.141350   0.017633  -8.016 1.09e-15 ***
nprenatal   -0.029895   0.011074  -2.699 0.006945 **
```

9

9



10

10

Covariate balance

- Propensity scores can be used to balance the treatment and control groups overall in a few ways:
  - Matching (e.g., matching treated cases to controls with same/close p-scores)
  - Weighting (e.g., applying inverse probability of treatment weights to the controls to make their distribution look like the treatment group)
- Balance on the propensity score, however, does not guarantee that the treatment and control groups will be balanced on each of the elements that go into the propensity score

11

11

Assessing balance with standardized difference

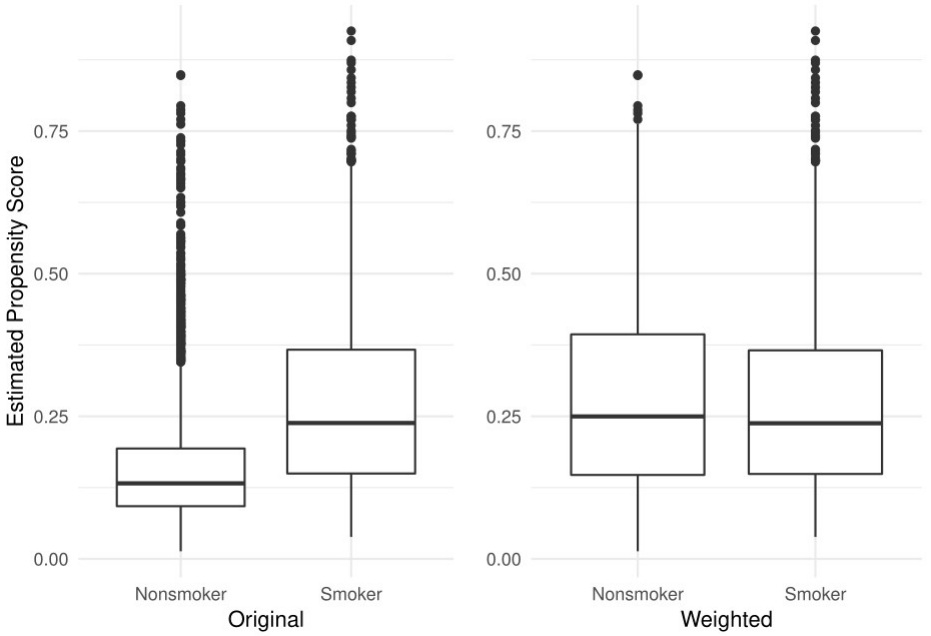
$$bias = \frac{\bar{x}_{treated} - \bar{x}_{control}}{\sqrt{\frac{s^2_{treated} + s^2_{control}}{2}}}$$

This is what psychologists call an "effect size" – the difference in z-scores

The traditional rule of thumb is that this should be no greater than .1 for any T&C comparison in any stratum for any variable.

12

12



13

13

```r
# data prep
dd <- as.data.frame(d)                # twang doesn't like tibbles
cov_names <-                          # list of covariate names
  c("mmarried", "alcohol", "mrace", "fbaby", "mage", "medu", "nprenatal")

# using twang::bal.stat
balance_table <-
  bal.stat(dd,
           vars = cov_names,
           treat.var = "mbsmoke",
           w.all = dd$attwt ,         # put the weight here
           sampw = 1 ,                # survey weights if needed (or 1)
           estimand = "ATT",          # which SD for std comparison (T vs. pooled)
           get.ks = FALSE ,           # don't need KS stats
           multinom = FALSE )         # set FALSE for binary treatment
balance_table$results %>% round(.,3)  # look at results and round to 3 decimals
```

Don't pay attention to the tests ("stat" and "p"). They are not really appropriate. Just keep an eye on the standardized differences.

	tx.mn <dbl>	tx.sd <dbl>	ct.mn <dbl>	ct.sd <dbl>	std.eff.sz <dbl>	stat <dbl>	p <dbl>
mmarried	0.473	0.500	0.470	0.499	0.008	0.178	0.859
alcohol	0.091	0.288	0.096	0.294	-0.014	-0.238	0.812
mrace	0.809	0.393	0.807	0.394	0.004	0.088	0.930
fbaby	0.372	0.483	0.363	0.481	0.018	0.435	0.664
mage	25.167	5.301	25.083	5.342	0.016	0.374	0.708
medu	11.639	2.168	11.330	3.486	0.142	1.848	0.065
nprenatal	9.862	4.208	9.766	4.066	0.023	0.488	0.625

14

14

Overview of options

- **Common support**: do we need to drop off-support cases?*
- **Distance metric**: match on p-score or logit of p-score?
- **Caliper**: how far is the "nearest neighbor" allowed to be?
- **Replacement**: should controls be allowed to be reused?
- **Ratio**: 1-to-1, k-to-1, or variable ratio matching?
- **Exact matching**: should we match exactly on one or more categorical variables?

*Already covered; same issues apply

15

15

Covariate Balancing Propensity Scores (CBPS)

- This is a newer technique that jointly maximizes the balance of the covariates and the prediction of the treatment using an empirical likelihood approach
- For full info, see Imai and Ratkovic (2014) and some newer developments in Fan et al. (https://imai.fas.harvard.edu/research/CBPStheory.html)

16

16

The intuition

- Conditioning on the true propensity score would satisfy the CIA; unfortunately we never have it
- If the propensity model is misspecified (as it almost always is), covariate imbalance (and thus bias) can result
- Optimizing covariate balance directly reduces this danger
- The CBPS estimator minimizes imbalance and maximizes prediction of treatment selection simultaneously

17

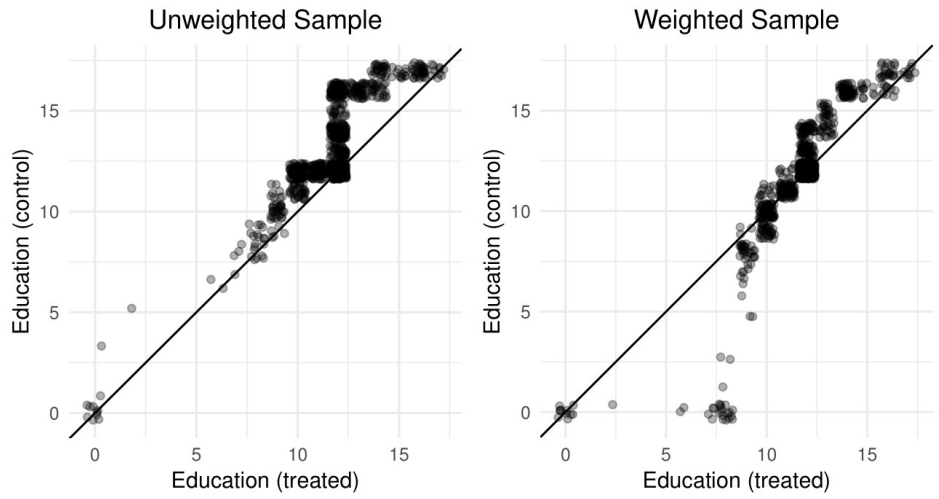
17

	tx.mn <dbl>	tx.sd <dbl>	ct.mn <dbl>	ct.sd <dbl>	std.eff.sz <dbl>	stat <dbl>	p <dbl>	ks <dbl>	ks.pval <dbl>
mmarried	0.473	0.500	0.473	0.499	0.002	0.043	0.966	0.001	1.000
alcohol	0.091	0.288	0.092	0.289	-0.001	-0.014	0.989	0.000	1.000
mrace	0.809	0.393	0.809	0.393	0.000	0.003	0.998	0.000	1.000
fbaby	0.372	0.483	0.371	0.483	0.001	0.025	0.980	0.000	1.000
mage	25.167	5.301	25.159	6.027	0.001	0.033	0.974	0.058	0.038
medu	11.639	2.168	11.636	3.167	0.001	0.023	0.981	0.087	0.000
nprenatal	9.862	4.208	9.859	4.031	0.001	0.017	0.987	0.019	0.984

As expected, overall and covariate balance are *very* good. I add the KS statistics here, however, just to show that even though the first moments (i.e., means) are very well balanced, the distributions are not equivalent.

18

18



These QQ plots show that the weighted selection of controls is overrepresented at both tails (more 0s and more college educated) even though the *mean* years is balanced.

19

19

IPTW with these p-scores

```
# compare using svyglm
svyglm(zweight ~ mbsmoke ,
         design = svydesign(id = ~1,
                            data = dd ,
                            weights = dd$attwt.cb)) %>% tidy()
```

term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
(Intercept)	0.00825875	0.02643909	0.3123689	7.547742e-01
mbsmoke	-0.39528806	0.04224721	-9.3565473	1.250252e-20

20

20