Observational Studies and Propensity Scores

STA 320
Design and Analysis of Causal Studies
Dr. Kari Lock Morgan and Dr. Fan Li
Department of Statistical Science
Duke University

Makeup Class

- Rather than making you come in for class on a weekend to make up the lost classes, please read this paper:
- Rubin (2007). The design *versus* the analysis of observational studies for causal effects: Parallels with the design of randomized trials, *Statistics in Medicine*, **26**(1): 20-36.

Randomized Experiments

- Recent article about randomized experiments, and the controversy of not using them (if interested):

Method of study is criticized in group's health policy tests (NY Times, 2/3/14)

Randomized Experiments

- Randomized experiments are the gold standard for estimating causal effects
- However, often they are not feasible
  - Cost
  - Ethical reasons
  - Time
  - Compliance
- Often, we must use data from observational studies to estimate causal effects

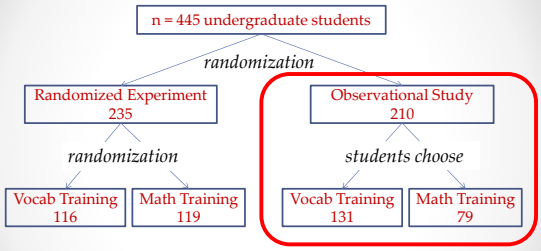
Observational Studies

- Say goodbye to unbiased estimates and exact p-values…
- … but if we only have observational data, we want to do the best we can!
- **GOAL: Make data from observational studies look as much as possible like data from a randomized experiment**

Timing of Treatment

- In a randomized experiment, the timing of the treatment assignment is clear => covariates are clear
- In an observational study, this may not be clear
- Solution? Clearly define the timing of the treatment (usually time of decision)

Shadish Data



Shadish, M. R., Clark, M. H., Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *JASA.* **103**(484): 1334-1344.

Design vs Analysis

- In a randomized experiment, the design phase (collecting data, balancing covariates, specifying plan) is done before access to outcomes and analysis
- In an observational study, you typically get all data together (covariates, treatment, outcomes): design and analysis mingled
- Solution? do the design part *before access to outcomes*

NO OUTCOMES!!!

- In a randomized experiment, outcomes are not available in the design phase
- When analyzing observational studies, the design should NOT include outcomes
- Anything done to try to make the data look "balanced" between treatment groups should IGNORE THE OUTCOMES!!!
- This is the only way to be objective
- As long as outcomes are hidden, you can do whatever you want to achieve covariate balance

Shadish Data

- Before doing anything else with the data (before even looking at it!):

REMOVE OUTCOMES!!!!!!!!!

"Design Trumps Analysis"

- "Design": everything done before access to outcome data
  - Contemplating and collecting data (including rich covariates)
  - Creating covariate balance
  - Specification of analysis plan once outcomes are revealed
- Idea: do all the hard work in the design phase, and the analysis with outcomes will be straightforward and easy

Analysis

- In randomized experiments, there is usually a pre-specified protocol for analysis
- In observational studies, people often try many different models and analyses – introduces subjectivity and bias
- Solution? specify protocol with outcomes in advance, and do most of your work in the design phase to make analysis easy

Assignment Mechanism

- In a randomized experiment, the assignment mechanism is regular (unconfounded, individualistic, probabilistic) by design
- In an observational study these are only assumptions, and may not hold
- Solution? Do what we can to make these assumptions more plausible

Unconfoundedness

- Based on the covariates, is the assignment independent of the outcomes?
- Why important? What if assignment did depend on the outcomes, conditional on the covariates?
- If not unconfounded, potential outcomes could differ between groups *before treatment even applied,* even if covariate values all the same!
- Unconfoundedness allows us to compare units with similar covariate values to estimate causal effects
- *Crucial assumption for causal inference*

Unconfoundedness

- The plausibility of unconfoundedness lies in the collection of rich covariates
- Want to compare "like with like".
- How many / what covariates do we need data on to ensure that a set of units are comparable?
- **Goal: data on all covariates that matter**
- Do we have all the important covariates?

Shadish Data - Covariates

- Vocab and math pre-test scores
- Number of math classes taken
- How much do you like math?
- How much do you like literature?
- Do you prefer math or literature?
- ACT score
- College GPA
- Age
- Gender
- (more, but we'll use these 10 here)

Unconfounded?

Reality

- In reality, observational studies are rarely truly unconfounded
- We just try to get as close as possible to the truth, by collecting the best covariate data possible
- (and using the techniques we'll learn…)

Probabilistic?

- Probabilistic: Every unit has some chance of being assigned to either treatment group, conditional on covariates
- Solution? If certain types of units are only observed in one group, eliminate these units (restrict causal inferences to units who might get either treatment)
- Remove units not similar to any units in the opposite group (often measured based on propensity score)

Assignment Mechanism

- In a randomized experiment, the assignment mechanism is known
- In an observational study the assignment mechanism is unknown
- Solution? Estimate the assignment mechanism by modeling it (propensity scores)

Regular Assignment Mechanism

- regular assignment mechanism:
  - unconfounded $\mathbf{W} \mid \mathbf{X}, \mathbf{Y} = \mathbf{W} \mid \mathbf{X}$
  - individualistic $W_i \mid \mathbf{X}, \mathbf{Y} = W_i \mid X_i, Y_i$
  - probabilistic $0 < p(W_i = 1 \mid \mathbf{X}, \mathbf{Y}) < 1$

=> $p(W_i = 1 \mid \mathbf{X}, \mathbf{Y}(1), \mathbf{Y}(0)) = p(W_i \mid X_i)$

- The probability a unit is treated depends only on that unit's covariates

Propensity Score

- For a regular assignment mechanism, the propensity score, e(x), is the probability of being in the treatment group, for X = x:

$$e(x) = P(W = 1 \mid X = x)$$

- Propensity scores are central to estimating causal effects from observational studies
- Unconfoundedness is *essential* for estimating causal effects

Propensity Score

- e(x) = P(W = 1 | X = x)
- One way to model it: logistic regression

$$\log\left(\frac{e(x)}{1 - e(x)}\right) = a + b'x$$

Propensity Scores in R

```
> ps.model = glm(W~.,data=shadish,family="binomial")
> summary(ps.model)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.0033176  1.3928412  -0.002   0.9981
vocabpre     0.0644999  0.0473549   1.362   0.1732
mathpre     -0.0146762  0.0796230  -0.184   0.8538
numbmath    -0.0478419  0.1578780  -0.303   0.7619
likemath    -0.1692489  0.1008835  -1.678   0.0934 .
likelit     -0.0506003  0.0944067  -0.536   0.5920
preflit      0.5318437  0.3040978   1.749   0.0803 .
actcomp     -0.0005051  0.0555204  -0.009   0.9927
collgpaa    -0.2012186  0.2552284  -0.788   0.4305
age         -0.0025896  0.0323975  -0.080   0.9363
male        -0.1311330  0.4133749  -0.317   0.7511
```

Is this a "good" model?

Estimating Propensity Scores

- GOAL: COVARIATE BALANCE
- Goal is NOT to get best estimates of propensity scores or best fitting model
- Estimating propensity scores involves fitting model, checking balance, and choosing the model which gives the best balance
- (however, we won't learn how to balance with propensity scores until next week)
- Important: NO OUTCOMES!

Estimating Propensity Scores

- Which covariates to include?
- The important ones! (substantively, not statistically)
- Variables to be included depend on subject matter knowledge (consult with subject expert), NOT statistical properties
- Variable selection? NO, not for primary covariates. Keep all important covariates in the model, even if insignificant.

Estimating Propensity Scores

- What about other covariates which we don't know *a priori* to be important, but might differ between treatment groups
  - other covariates recorded?
  - interactions?
  - functions of the covariates? (log, quadratic, ..)
- Variable selection… any way you know how
- likelihood ratios? stepwise regression?

Likelihood Ratio

- Add potential new variable into the model (one at a time), and compute LR statistic:

$$LR = -2\log\left(\frac{\text{likelihood for smaller model}}{\text{likelihood for larger model}}\right) \approx c^2_{df2-df1}$$

- If statistic exceeds a certain threshold, include variable. If not, don't include.
- Thresholds may differ for covariates, interactions, and functions

Likelihood Ratio in R

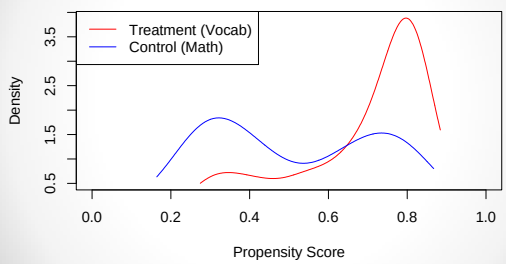
```
> ps.model = glm(W~.,data=shadish,family="binomial")
> ps.model2 = glm(W~.+I(age^2),data=shadish, family="binomial")
> lrtest(ps.model, ps.model2)

Likelihood ratio test

Model 1: W ~ vocabpre + mathpre + numbmath + likemath + likelit
+ preflit +
    actcomp + collgpaa + age + male
Model 2: W ~ vocabpre + mathpre + numbmath + likemath + likelit
+ preflit +
    actcomp + collgpaa + age + male + I(age^2)
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  11 -112.83
2  12 -112.75  1 0.1675     0.6823
```

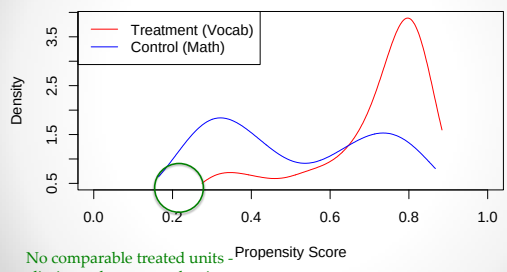
Propensity Scores

```
ps = predict(ps.model, type="response")
```



Probabilistic?

```
ps = predict(ps.model, type="response")
```



No comparable treated units - eliminate these control units

Estimating Propensity Scores

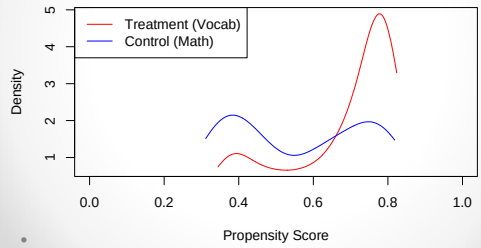
- In practice, estimating the propensity score is an iterative process:

1. Estimate propensity score
2. Eliminate units with no overlap (eliminate units with no comparable units in other groups)
3. Repeat until "probabilistic" – any unit could get treatment or control, based on covariates

Go back and refit ps model

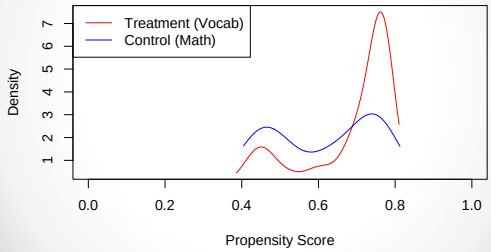
Estimating Propensity Score

```
shadish= shadish[ps>=min(ps[shadish$W==1]) &
          ps<=max(ps[shadish$W==0]),]
ps.model = glm(W~.,data=shadish, family="binomial")
summary(ps.model)
ps = predict(ps.model, type="response")
```



Estimating Propensity Score

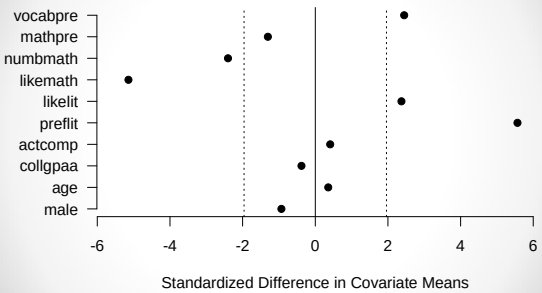
- Repeat, eliminate non-comparable units and refit propensity score model…



Covariate Balance

- In randomized experiments, the randomization creates covariate balance between treatment groups
- In observational studies, treatment groups will be naturally unbalanced regarding covariates
- Solution? compare similar units
- (How? Propensity score methods.)

Shadish Covariate Balance



GOAL: Try to fix this!

Subgroups

- If we have enough covariates (unconfounded), within subgroups of people with identical covariates, observational studies look like randomized experiments
- Idea: subclassify people based on similar covariate values, and estimate treatment effect within each subclass
- (similar to stratified experiments)

Comparison of Mortality Rates for Two Smoking Treatments in U.S.

	Cigarette Smokers	Cigar/Pipe Smokers
Mortality Rate per 1000 person-years, %	13.5	17.4

Cochran WG. The Effectiveness of Adjustment of Subclassification in Removing Bias in Observational Studies. Biometrics 1968; 24: 295-313.

Slide by Cassandra Pattanayak

One Key Covariate Smoking, Cochran (1968)

Population: Male smokers in U.S.

Active treatment: Cigar/pipe smoking

Control treatment: Cigarette smoking

Outcome: Death in a given year

Decision-Maker: Individual male smoker

Reason for smoking male to choose cigarettes versus cigar/pipe?

Age is a key covariate for selection of smoking type for males

Slide by Cassandra Pattanayak

Subclassification to Balance Age

To achieve balance on age, compare:

- "young" cigar/pipe smokers with "young" cigarette smokers
- "old" cigar/pipe smokers with "old" cigarette smokers

Better: young, middle-aged, old, or more age subclasses

Objective of study design, without access to outcome data: approximate a completely randomized experiment within each subclass

Only after finalizing design, reveal outcome data

Rubin DB. The Design Versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials. Statistics in Medicine, 2007.

Slide by Cassandra Pattanayak

Comparison of Mortality Rates for Two Smoking Treatments in U.S.

	Cigarette Smokers	Cigar/Pipe Smokers
Mortality Rate per 1000 person-years, %	13.5	17.4
Averaging Over Age Subclasses		
2 Age Subclasses	16.4	14.9
3 Age Subclasses	17.7	14.2
11 Age Subclasses	21.2	13.7

Cochran WG. The Effectiveness of Adjustment of Subclassification in Removing Bias in Observational Studies. Biometrics 1968; 24: 295-313.

Slide by Cassandra Pattanayak

This is a Simpson's Paradox!

(at least, approximately)

Slide by Cassandra Pattanayak

What if we had 20 covariates, with 4 levels each?

Over a million million subclasses

Slide by Cassandra Pattanayak

Solution?

- How can we balance many covariates?

BALANCE THE PROPENSITY SCORE!

Propensity Score

- Amazing fact: balancing on just the propensity score balances ALL COVARIATES included in the propensity score model!!!

- (this is the topic for next week)
- How do we compare units with similar propensity scores?
  - Subclassification
  - Matching
  - Weighting

Select Facts about Classical Randomized Experiments

Timing of treatment assignment clear

Design and Analysis separate by definition: design automatically "prospective," without outcome data

Unconfoundedness, probabilisticness by definition

Assignment mechanism – and so propensity scores – known

Randomization of treatment assignment leads to expected balance on covariates
("Expected Balance" means that the joint distribution of covariates is the same in the active treatment and control groups, on average)

Analysis defined by protocol rather than exploration

Slide by Cassandra Pattanayak

Select Facts about Observational Studies

Timing of treatment assignment may not be specified

Separation between design and analysis may become obscured, if covariates and outcomes arrive in one data set

Unconfoundedness, probabilisticness not guaranteed

Assignment mechanism – and therefore propensity scores – unknown

Lack of randomization of treatment assignment leads to imbalances on covariates

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

46

Best Practices for Observational Studies

Timing of treatment assignment may not be specified

Separation between design and analysis may become obscured, if covariates and outcomes arrive in one data set

Unconfoundedness, probabilisticness not guaranteed

Assignment mechanism – and therefore propensity scores – unknown

Lack of randomization of treatment assignment leads to imbalances on covariates

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

47

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

Separation between design and analysis may become obscured, if covariates and outcomes arrive in one data set

Unconfoundedness, probabilisticness not guaranteed

Assignment mechanism – and therefore propensity scores – unknown

Lack of randomization of treatment assignment leads to imbalances on covariates

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

48

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

2. Hide outcome data until design phase is complete

Unconfoundedness, probabilisticness not guaranteed

Assignment mechanism – and therefore propensity scores – unknown

Lack of randomization of treatment assignment leads to imbalances on covariates

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

2. Hide outcome data until design phase is complete

3. Identify key covariates likely related to outcomes and/or treatment assignment. If key covariates not observed or very noisy, usually better to give up and find a better data source.

4. Remove units not similar to any units in opposite treatment group

Assignment mechanism – and therefore propensity scores – unknown

Lack of randomization of treatment assignment leads to imbalances on covariates

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

2. Hide outcome data until design phase is complete

3. Identify key covariates likely related to outcomes and/or treatment assignment. If key covariates not observed or very noisy, usually better to give up and find a better data source.

4. Remove units not similar to any units in opposite treatment group

Assignment mechanism – and therefore propensity scores – unknown

Lack of randomization of treatment assignment leads to imbalances on covariates

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

2. Hide outcome data until design phase is complete

3. Identify key covariates likely related to outcomes and/or treatment assignment. If key covariates not observed or very noisy, usually better to give up and find a better data source.

4. Remove units not similar to any units in opposite treatment group

5. Estimate propensity scores, as a way to…

6. Find subgroups (subclasses or pairs) in which the active treatment and control groups are balanced on covariates (not always possible; inferences limited to subgroups where balance is achieved)

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

2. Hide outcome data until design phase is complete

3. Identify key covariates likely related to outcomes and/or treatment assignment. If key covariates not observed or very noisy, usually better to give up and find a better data source.

4. Remove units not similar to any units in opposite treatment group

5. Estimate propensity scores, as a way to…

6. Find subgroups (subclasses or pairs) in which the treatment groups are balanced on covariates (not always possible; inferences limited to subgroups where balance is achieved)

Analysis often exploratory rather than defined by protocol

Slide by Cassandra

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

2. Hide outcome data until design phase is complete

3. Identify key covariates likely related to outcomes and/or treatment assignment. If key covariates not observed or very noisy, usually better to give up and find a better data source.

4. Remove units not similar to any units in opposite treatment group

5. Estimate propensity scores, as a way to…

6. Find subgroups (subclasses or pairs) in which the active treatment and control groups are balanced on covariates (not always possible; inferences limited to subgroups where balance is achieved)

7. Analyze according to pre-specified protocol

Slide by Cassandra

Best Practices for Observational Studies

1. Determine timing of treatment assignment relative to measured variables

Design Observational Study to Approximate Hypothetical, Parallel Randomized Experiment

5. Estimate propensity scores, as a way to…

6. Find subgroups (subclasses or pairs) in which the active treatment and control groups are balanced on covariates (not always possible; inferences limited to subgroups where balance is achieved)

7. Analyze according to pre-specified protocol

Slide by Cassandra

55

To Do

- Read Rubin 2007
- Read Ch 12, 13
- Homework 3 (due Monday)
- Bring laptops to class on Monday