Understanding Propensity Score Matching

Ehsan Karim

2021-06-02

Contents

1 Description **5**
- 1.1 Prerequisites . . . 5
- 1.2 Main references . . . 5

2 Useful Terminologies **7**
- 2.1 Potential outcome . . . 7
- 2.2 Parameters of interest . . . 7
- 2.3 Balance . . . 11
- 2.4 Adjustment . . . 13
- 2.5 Overlap . . . 14

3 Propensity score **15**
- 3.1 Motivating problem . . . 15
- 3.2 Defining Propensity score . . . 18
- 3.3 PS Matching Steps . . . 19

4 Step 1: Exposure modelling **21**
- 4.1 Model specification . . . 21
- 4.2 Variables to adjust . . . 22
- 4.3 Model selection . . . 23
- 4.4 Alternative modelling strategies . . . 23
- 4.5 PS estimation . . . 24

5 Step 2: Propensity score Matching **25**
- 5.1 Matching method NN . . . 25
- 5.2 Initial fit . . . 25
- 5.3 Fine tuning: add caliper . . . 26
- 5.4 Matches . . . 27
- 5.5 Other matching algorithms . . . 27

6 Step 3: Balance and overlap **29**
- 6.1 Assessment of Balance by SMD . . . 29
- 6.2 Vizualization for Overlap . . . 30
- 6.3 SMD vs. P-values . . . 32

6.4 Variance ratio 32
6.5 Close inspection of boundaries 33
6.6 Unsatirfactory balance 34

7 Step 4: Outcome modelling **35**
7.1 Crude outcome model 35
7.2 Double-adjustment 35
7.3 Adjusted outcome model 36
7.4 Other cosiderations for outcome model 36
7.5 Estimate obtained 37

8 PS vs. Regression **39**
8.1 Data Simulation 39
8.2 Treatment effect from counterfactuals 41
8.3 Treatment effect from Regression 42
8.4 Treatment effect from PS 42
8.5 Non-linear Model 43
8.6 Common misconception 47
8.7 Benifits of PS 47
8.8 Limitations of PS 48

9 Reporting Guidelines **49**
9.1 Discipline-specific Reviews 49
9.2 Suggested Guidelines 49
9.3 Software 50
9.4 Further Resources 50

References **51**

Chapter 1

Description

Propensity score matching is widely used in analyzing observational datasets to reduce the impact of confounding due to observed covariates. This workshop will provide a basic overview of related causal inference concepts, explain propensity score matching analysis steps, illustrate propensity score matching diagnostics, and provide examples of when this method may be preferable to a regression.

1.1 Prerequisites

The prerequisites are knowledge of multiple regression analysis and basic probability. Software demonstrations and codes will be provided in R, although proficiency in R is not required for understanding the concepts.

1.2 Main references

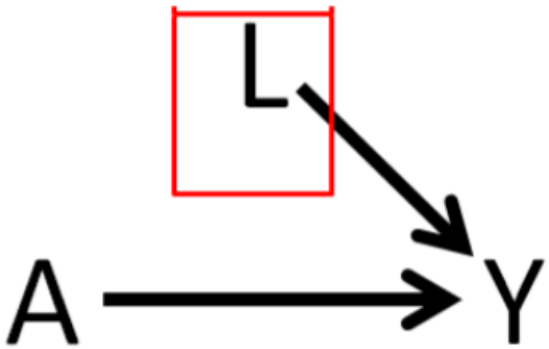
- Propensity score matching steps Austin (2011b)
- Reporting guideline Karim et al. (2020)

Chapter 2

Useful Terminologies

2.1 Potential outcome

- A: Exposure status
  - 1 = takes Rosuvastatin
  - 0 = does not take rosuvastatin
- Y: Outcome: Total cholesterol levels
  - $Y(A = 1)$ = potential outcome when exposed
  - $Y(A = 0)$ = potential outcome when not exposed
- L: Confounder: Age



2.2 Parameters of interest

When assessing the effect of an exposure on an outcome, we are interested about the following estimands

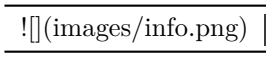
- treatment effect for an individual (TE)
- average treatment effect (ATE)

- average treatment effect on the treated (ATT)

2.2.1 TE

- John takes Rosuvastatin ($A = 1$) and his total cholesterol level is $= Y(A = 1) = 195$ mg/dL (milligrams per deciliter) after 3 months
- John does not take Rosuvastatin ($A = 0$) and his total cholesterol level is $= Y(A = 0) = 245$ mg/dL after 3 months Effect of Rosuvastatin on John is $=$

$TE = Y(A = 1) - Y(A = 0) = 195 - 245 = -50$



2.2.2 ATE

```
Person <- c("John","Jim","Jake","Cody","Luke")
Y1 <- c( 195, 100, 210, 155, 165)
Y0 <- c(245, 160, 270, 210, 230)
PotentialOutcomes <- data.frame(Person, Y1, Y0, TE = Y1-Y0)
mean.values <- c(NA, mean(PotentialOutcomes$Y1),
                 mean(PotentialOutcomes$Y0),
                 mean(PotentialOutcomes$TE))
PotentialOutcomes <- rbind(PotentialOutcomes, mean.values)
kable(PotentialOutcomes, booktabs = TRUE,
             col.names = c("Person", "Y(1)", "Y(0)", "TE")) %>%
  row_spec(6, bold = T, color = "white", background = "#D7261E")
```

Person	Y(1)	Y(0)	TE
John	195	245	-50
Jim	100	160	-60
Jake	210	270	-60
Cody	155	210	-55
Luke	165	230	-65
	165	**223**	**-58**

$ATE = E[Y(A = 1) - Y(A = 0)]$

```
mean(PotentialOutcomes$Y1 - PotentialOutcomes$Y0)
```

```
## [1] -58
```

2.2.3 Interpretation of ATE

This is a treatment effect (on an average) of the following hypothetical situation

- having the entire population as treated, vs.

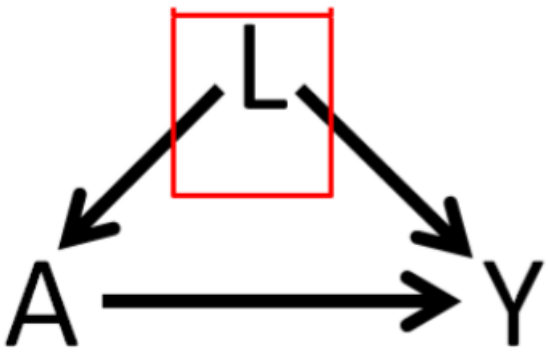
- having the entire population as untreated.

Entire population is the reference goup here.

2.2.4 Identifiability Assumptions

If we can compute a causal quantity, such as $ATE = E[Y(A = 1) - Y(A = 0)]$ using a statistical quantity, such as `mean(PotentialOutcomes$Y1 - PotentialOutcomes$Y0)`, we say that the causal quantity is identifiable.

Exchangeability	$Y(1), Y(0) \perp A$	Treatment assignment is independent of the potential outcome
Positivity	$0 < P(A = 1) < 1$	Subjects are eligible to receive both treatment
Consistency	$Y = Y(a) \forall A = a$	No multiple version of the treatment
No interference		Treated one patient will not impact outcome for others



Extending these assumptions when confounders exist:

Conditional Exchangeability	$Y(1), Y(0) \perp A \mid L$	Treatment assignment is independent of the potential outcome, given L
Positivity	$0 < P(A = 1 \mid L) < 1$	Subjects are eligible to receive both treatment, given L

2.2.5 ATT

- Assume that the following are the confounders that impact the relationship between rosuvastatin and cholesterol levels
  - race
  - sex
  - age
- We have 5 Rosuvastatin-treated subjects who are all
  - white,
  - male,
  - 50 years of age
- We recruited additional 5 subjects (same characteristics) to non-rosuvastatin group.

Treated group:

```
Person <- c("John","Jim","Jake","Cody","Luke")
Y1 <- c( 195, 100, 210, 155, 165)
Y0 <- rep(NA, length(Y1))
Treated <- data.frame(Person, Y1, Y0, TE = Y1-Y0)
Treated[6,2] <- mean(Treated$Y1)
kable(Treated, booktabs = TRUE,
              col.names = c("Person", "Y(1)", "Y(0)", "TE"))%>%
  row_spec(6, bold = T, color = "white", background = "#D7261E")
```

Person	Y(1)	Y(0)	TE
John	195		
Jim	100		
Jake	210		
Cody	155		
Luke	**165**		
	165		

Untreated group: New folks with characteristics similar to the treated group.

```
Person <- c( "Jack", "Dustin", "Cole", "Lucas", "Dylan")
Y0 <- c( 245, 160, 270, 210, 165)
Y1 <- rep(NA, length(Y0))
Untreated <- data.frame(Person, Y1, Y0, TE = Y1-Y0)
Untreated[6,3] <- mean(Untreated$Y0)
kable(Untreated, booktabs = TRUE,
              col.names = c("Person", "Y(1)", "Y(0)", "TE"))%>%
  row_spec(6, bold = T, color = "white", background = "#D7261E")
```

Person	Y(1)	Y(0)	TE
Jack		245	
Dustin		160	
Cole		270	
Lucas		210	
Dylan		165	
		210	

$ATT = E[Y(A = 1) - Y(A = 0)|A = 1]$

```
mean(Treated$Y1) - mean(Untreated$Y0)
```

```
## [1] -45
```

2.2.6 Interpretation of ATT

This is a treatment effect (on an average) of

- the treated population (reference group), vs.
- untreated population, but have similar characteristics to the reference group/treated population.

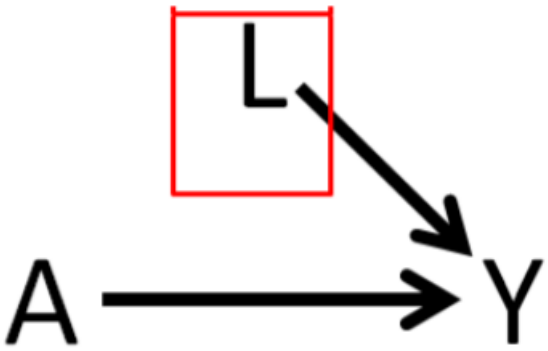
It is also possible to change the reference population to untreated population. Then it is called Average Treatment Effect for the Untreated (ATU).

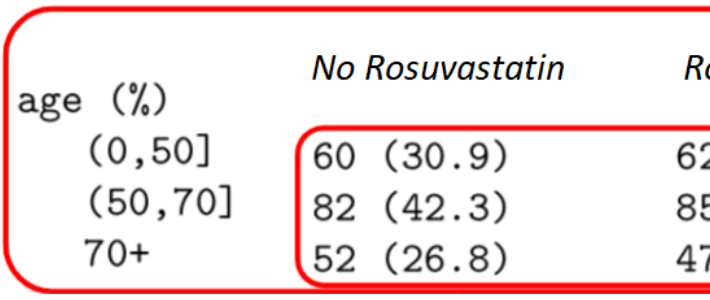
2.2.7 ATT vs. ATE

	In a RCT (enough n), the ATT & ATE are equivalent
	In an observational study the ATT and ATE are not necessarily the same.

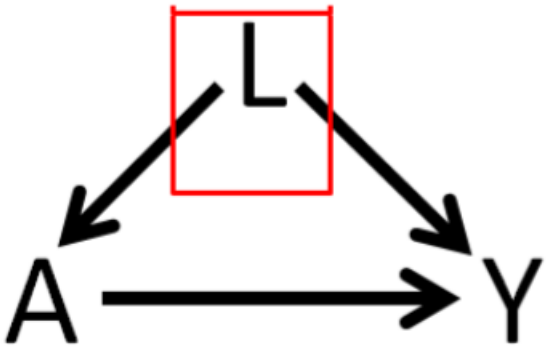
2.3 Balance

Balance in RCT:





In absence of randomization:



age	No Rosuvastat
(0,50]	0.63
(50,70]	0.26
70+	0.11

2.3.1 Measures of Balance

2.3.1.1 SMD

Austin (2011a)

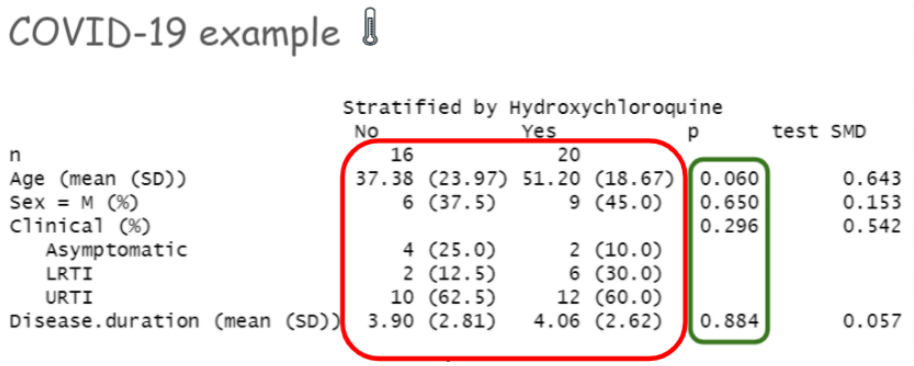
- For continuous confounders:
  - $SDM_{continuous} = \frac{\bar{L}_{Rosuvastatin} - \bar{L}_{NoRosuvastatin}}{\sqrt{\frac{s^2_{Rosuvastatin} + s^2_{NoRosuvastatin}}{2}}}$
- For binary confounders:
  - $SDM_{binary} = \frac{\hat{p}_{Rosuvastatin} - \hat{p}_{NoRosuvastatin}}{\sqrt{\frac{\hat{p}_{Rosuvastatin} \times (1-\hat{p}_{Rosuvastatin}) + \hat{p}_{NoRosuvastatin} \times (1-\hat{p}_{NoRosuvastatin})}{2}}}$

	No Rosuvastatin	Rosuvastatin	SMD
n	4089	325	
gender = Female (%)	1960 (47.9)	194 (59.7)	0.238
diabetes = Yes (%)	358 (8.8)	87 (26.8)	0.485
smoke = Yes (%)	1796 (43.9)	177 (54.5)	0.212
age (%)			0.891
(0,50]	2577 (63.0)	74 (22.8)	
(50,70]	1046 (25.6)	169 (52.0)	
70+	466 (11.4)	82 (25.2)	

Generally, 0.1 is used as a cut-point. But some suggest more liberal cut-points. More on that later.

COVID example from Gautret et al. (2020)

p-value vs. SMD



2.3.1.2 Variance ratio

Variances of baseline characteristics between comparator groups under consideration. Suggested cut-point rages are (0.5 to 2). More liberal cutpoints are also used in the literature. More on this later.

2.4 Adjustment

2.4.1 Why adjust?

In absence of randomization, treatment effect estimate ATE $= E[Y|A = 1] - E[Y|A = 0]$ includes

- Treatment effect
- Systematic differences in 2 groups ('confounding')
  - Doctors may prescribe tx more to frail and older age patients.
  - In here, $L =$ age is a confounder.

In absence of randomization, if age is a known confounder:

Causal effect for young (< 50)	$E[Y\|A = 1, L =$ `younger age`$]$ - $E[Y\|A = 0, L =$ `younger age`$]$
Causal effect for old (≥ 50)	$E[Y\|A = 1, L =$ `older age`$]$ - $E[Y\|A = 0, L =$ `older age`$]$

Conditional exchangeability; only works if L is measured.

2.4.2 Adjustment Methods

Adjustment could mean

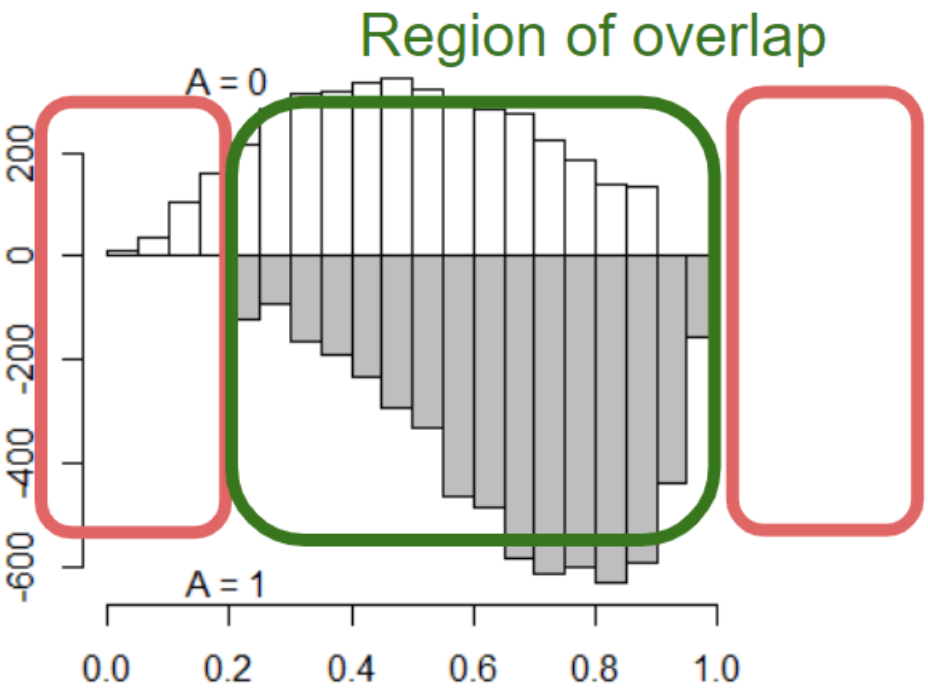
- exact matching
- stratification

 When L includes a large number of covariates, matching method would result in

Regression is also a popular adjustment method.

2.5 Overlap

- "Lack of complete overlap" happens if there is a baseline covariate space where there are exposed patients, but no control or vice versa.
  - Region of 'no overlap' is an inherent limitation of the data.



- **Regression adjustment** usually do not offer any solution to this.
  - Consequently, inference is not generalizable beyond the region of overlap.

Chapter 3

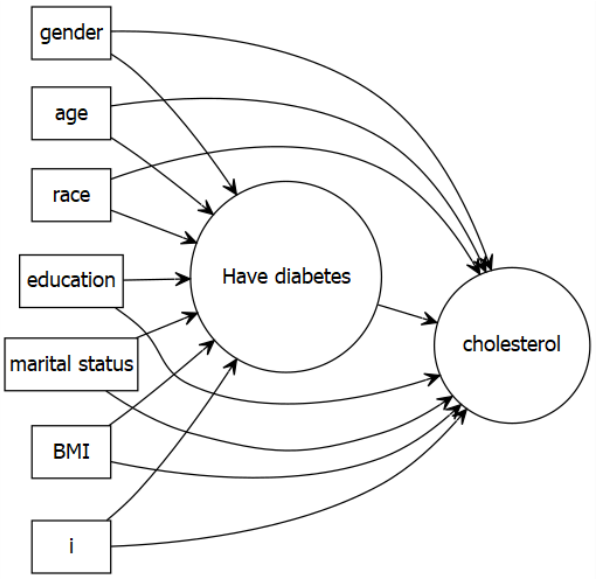
Propensity score

3.1 Motivating problem

Y : Outcome	Cholesterol levels (high vs. low)
A : Exposure	Diabetes
L : Known Confounders	gender, age, race, education, married, BMI

Search literature for the confounder variables, and look for those variables in the data source (NHANES 2017-2018).

```
load(file="data/NHANES17.RData")
require(dplyr)
analytic <- dplyr::select(analytic,
                  cholesterol, # outcome
                  gender, age, race, education,
                  married, bmi, # confounders
                  diabetes) # exposure
analytic$cholesterol <- ifelse(analytic$cholesterol > 240, 1, 0)
analytic$diabetes <- ifelse(analytic$diabetes == "Yes", 1, 0)
```




```
library(Hmisc)
```

```
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following object is masked from 'package:jtools':
##
##     %nin%
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
describe(analytic)
```

```
## analytic
##
##  8  Variables      1562  Observations
## ---------------------------------------------------------------------------
```

```
## cholesterol
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1562        0        2    0.292      171   0.1095   0.1951
##
## ------------------------------------------------------------------------------
## gender
##        n  missing distinct
##     1562        0        2
##
## Value      Female   Male
## Frequency     603    959
## Proportion  0.386  0.614
## ------------------------------------------------------------------------------
## age : Age in years at screening
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1562        0       61    0.999    53.18    19.78       25       29
##      .25      .50      .75      .90      .95
##       38       55       67       76       80
##
## lowest : 20 21 22 23 24, highest: 76 77 78 79 80
## ------------------------------------------------------------------------------
## race
##        n  missing distinct
##     1562        0        4
##
## Value         Black Hispanic    Other    White
## Frequency       324      284      228      726
## Proportion    0.207    0.182    0.146    0.465
## ------------------------------------------------------------------------------
## education
##        n  missing distinct
##     1562        0        3
##
## Value          College High.School      School
## Frequency          806         658          98
## Proportion       0.516       0.421       0.063
## ------------------------------------------------------------------------------
## married
##        n  missing distinct
##     1562        0        3
##
## Value                 Married      Never.married Previously.married
## Frequency                 921                228                413
## Proportion              0.590              0.146              0.264
## ------------------------------------------------------------------------------
## bmi : Body Mass Index (kg/m**2)
```

```
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##     1562        0      314        1    29.96    7.972    20.00    21.71
##      .25      .50      .75      .90      .95
##    25.00    28.90    33.80    39.59    43.69
##
## lowest : 14.8 15.1 15.5 15.7 16.2, highest: 57.2 60.3 61.6 61.9 64.2
## ------------------------------------------------------------------------------
## diabetes
##        n  missing distinct     Info      Sum     Mean      Gmd
##     1562        0        2      0.5      330   0.2113   0.3335
##
## ------------------------------------------------------------------------------
```

3.2 Defining Propensity score

- Conditional Probability of getting treatment, given the observed covariates

  Prob(treatment: $A = 1$ | baseline or pre-treatment covariates: L)

  Prob($A = 1$: Has diabetes | L: gender, age, race, education, married, bmi)

- PS = $Prob(A = 1|L)$

3.2.1 Theoretical result

Rosenbaum and Rubin (1983) showed:

- For potential outcomes $Y(1), Y(0)$, if you have sufficient observed covariate list L to reduce confounding ('strong ignoribility'):
  - i.e., if $(Y(1), Y(0)) \perp A|L$
  - Note that is this NOT $Y \perp A|L$
- then
  - $(Y(1), Y(0)) \perp A|PS$ and
  - $A \perp L|PS$

3.2.2 Assumptions

Conditional Exchangeability	$Y(1), Y(0) \perp A\|L$	Treatment assignment is independent of the potential outcome, given L
Positivity	$0 < P(A = 1\|L) < 1$	Subjects are eligible to receive both treatment, given L

Consistency	$Y = Y(a) \forall A = a$	No multiple version of the treatment

3.2.3 Ways to use PS

Many ways to use propensity scores (PS) in the analysis

- **PS matching** [our focus today: intuitive!]
- PS weighting
- PS stratification
- PS used as a covariate

3.3 PS Matching Steps

Propensity score matching has 4 steps (Austin, 2011b)

Step 1	exposure modelling: $PS = Prob(A = 1\|L)$
Step 2	Match by PS
Step 3	Assess balance and overlap (PS and L)
Step 4	outcome modelling: $Prob(Y = 1\|A = 1)$

Chapter 4

Step 1: Exposure modelling

4.1 Model specification

Specify the propensity score model to estimate propensity scores, and fit the model:

$A \sim L$

```
baselinevars <- c("gender", "age", "race", "education", "married", "bmi")
ps.formula <- as.formula(paste("diabetes", "~", paste(baselinevars, collapse = "+")))
ps.formula
```

```
## diabetes ~ gender + age + race + education + married + bmi
```

```
# fit logistic regression to estimate propensity scores
PS.fit <- glm(ps.formula,family="binomial", data=analytic)
require(jtools)
summ(PS.fit)
```

Observations	1562
Dependent variable	diabetes
Type	Generalized linear model
Family	binomial
Link	logit

$\chi^2(10)$	282.89
Pseudo-R^2 (Cragg-Uhler)	0.26
Pseudo-R^2 (McFadden)	0.18
AIC	1349.94
BIC	1408.83

	Est.	S.E.	z val.	p
(Intercept)	-8.38	0.58	-14.49	0.00
genderMale	0.34	0.15	2.26	0.02
age	0.06	0.01	11.26	0.00
raceHispanic	0.15	0.23	0.64	0.52
raceOther	0.76	0.23	3.25	0.00
raceWhite	-0.23	0.18	-1.23	0.22
educationHigh.School	0.14	0.15	0.95	0.34
educationSchool	0.52	0.27	1.92	0.05
marriedNever.married	-0.04	0.25	-0.16	0.88
marriedPreviously.married	-0.02	0.16	-0.15	0.88
bmi	0.10	0.01	10.14	0.00

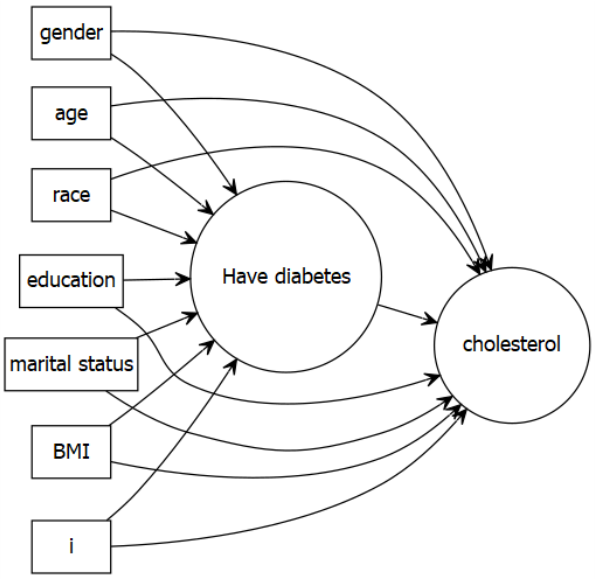
Standard errors: MLE

- Coef of PS model fit is not of concern
- Model can be rich: to the extent that prediction is better
- But look for multi-collinearity issues
  - SE too high?

4.2 Variables to adjust

Brookhart et al. (2006)

- Observed covariates are used to fix design
- Which covariates should be selected:
  - known to be a confounder (causes of Y and A)
  - known to be a cause of the outcome (risk factors of Y)
  - avoid known instruments or noise variables: **SE suffers**
  - mediating factors should be avoided (total effect = goal)
- Try drawing causal diagram to determine which variables to include



4.3 Model selection

Usually done for the variables that are *not known as a confounder* in the literature, or based on subject area knowledge.

- Stepwise (p-value or criterion based) not recommended
  - depending on sample size, different values can get selected
  - may select variables highly associated with A
- Don't look at the outcome (Y) in your data to select covariates
  - There are debate about this (ideal vs. pragmatism)
  - see Karim et al. (2018) for an example.

4.4 Alternative modelling strategies

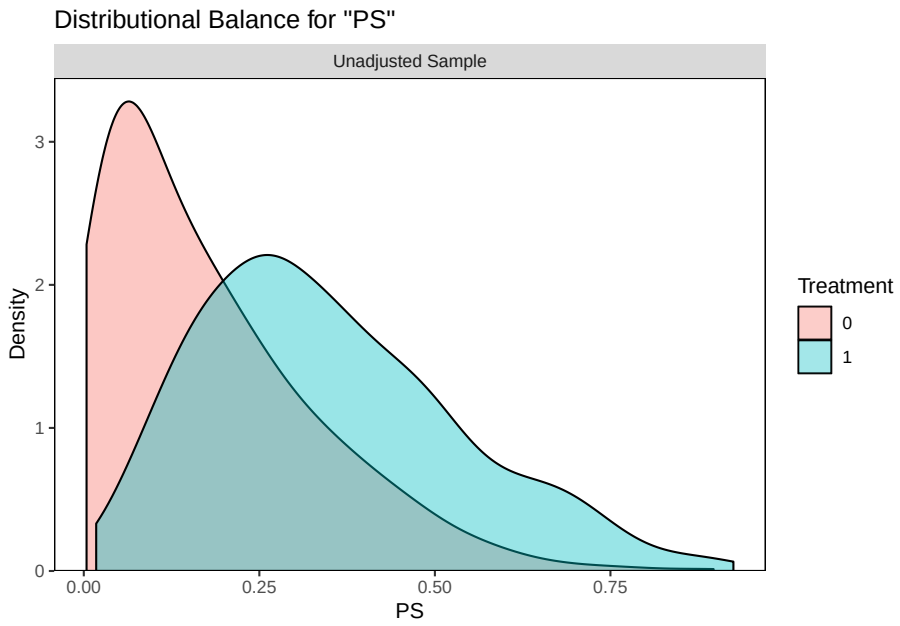
- Other machine learning alternatives are possible to use instead of logistic regression.
  - tree based methods have better ability to detect non-linearity / non-additivity (**model-specification** aspect)
  - shrinkage methods - lasso / elastic net may better deal with multi-collinearity
  - ensemble learners / super learners were successfully used
  - shallow/deep learning!

4.5 PS estimation

PS is unknown, and needs to be estimated from the fitted exposure model:

```
# extract estimated propensity scores from the fit
analytic$PS <- predict(PS.fit, newdata = analytic, type="response")
```

```
require(cobalt)
bal.plot(analytic, var.name = "PS",
         treat = "diabetes",
         which = "both",
         data = analytic)
```



	Don't loose sight that better **balance** is the ultimate goal for propensity sc
	Prediction of A is just a means to that end (as true PS is unknown)
	May attract variables highly associated with A

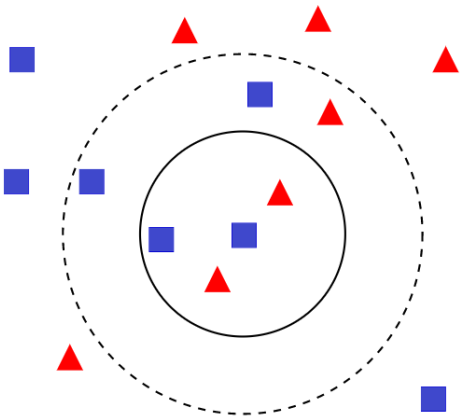
Chapter 5

Step 2: Propensity score Matching

5.1 Matching method NN

Match using estimates propensity scores

- nearest-neighbor (NN) matching
- without replacement
- with caliper = .2*SD of logit of propensity score
- with 1:1 ratio (pair-matching)



5.2 Initial fit

1:1 NN Match using estimates propensity scores

```
set.seed(123)
require(MatchIt)
match.obj <- matchit(ps.formula, data = analytic,
                     distance = 'logit',
                     method = "nearest",
                     replace=FALSE,
                     ratio = 1)
analytic$PS <- match.obj$distance
summary(match.obj$distance)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.003916 0.068128 0.169946 0.211268 0.312987 0.925132
```

```
match.obj
```

```
## A matchit object
##  - method: 1:1 nearest neighbor matching without replacement
##  - distance: Propensity score
##              - estimated with logistic regression
##  - number of obs.: 1562 (original), 660 (matched)
##  - target estimand: ATT
##  - covariates: gender, age, race, education, married, bmi
```

5.3 Fine tuning: add caliper

2 SD of logit of the propensity score is suggested as a caliper.

```
logitPS <-  -log(1/analytic$PS - 1)
# logit of the propensity score
.2*sd(logitPS) # suggested in the literature
```

```
## [1] 0.2606266
```

```
# choosing too strict PS has unintended consequences
```

```
set.seed(123)
require(MatchIt)
match.obj <- matchit(ps.formula, data = analytic,
                     distance = 'logit',
                     method = "nearest",
                     replace=FALSE,
                     caliper = .2*sd(logitPS),
                     ratio = 1)
analytic$PS <- match.obj$distance
summary(match.obj$distance)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
```

```
## 0.003916 0.068128 0.169946 0.211268 0.312987 0.925132
```

```
match.obj
```

```
## A matchit object
##  - method: 1:1 nearest neighbor matching without replacement
##  - distance: Propensity score [caliper]
##              - estimated with logistic regression
##  - caliper: <distance> (0.045)
##  - number of obs.: 1562 (original), 632 (matched)
##  - target estimand: ATT
##  - covariates: gender, age, race, education, married, bmi
```

5.4 Matches

Taking a closer look at the matches

```
# Ref: https://lists.gking.harvard.edu/pipermail/matchit/2013-October/000559.html
matches <- as.data.frame(match.obj$match.matrix)
colnames(matches)<-c("matched_unit")
matches$matched_unit<-as.numeric(
  as.character(matches$matched_unit))
matches$treated_unit<-as.numeric(rownames(matches))
matches.only<-matches[!is.na(matches$matched_unit),]
head(matches.only)
```

```
##     matched_unit treated_unit
## 40          8496           40
## 56          3139           56
## 65          4192           65
## 66            94           66
## 86          2212           86
## 110         7154          110
```

5.5 Other matching algorithms

More NN ratio is usually better. But creates issue when calculating variances (but can be easily handled).

Other possibilities

- Optimal
- genetic matching
- CEM
- variable ratio NN

Chapter 6

Step 3: Balance and overlap

Balance is more important than prediction!

- Criteria to assess success of step 2: PS estimation
  - better balance
  - better overlap [no extrapolation!]
  - PS = 0 or PS = 1 needs close inspection

6.1 Assessment of Balance by SMD

- balance = similarity of the covariate distributions
- d or $SMD > 0.1$ can be considered as imbalance (Austin, 2011a)

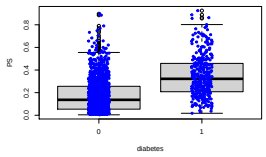
```
tab1e <- CreateTableOne(vars = baselinevars,
                        data = analytic,
                        strata = "diabetes",
                        includeNA = TRUE,
                        test = TRUE, smd = TRUE)
print(tab1e, showAllLevels = FALSE, smd = TRUE, test = TRUE)
```

```
##                         Stratified by diabetes
##                          0             1             p      test SMD
##   n                       1232           330
##   gender = Male (%)        738 (59.9)    221 (67.0)   0.023       0.147
##   age (mean (SD))        50.54 (17.23) 63.04 (12.87) <0.001       0.822
##   race (%)                                            0.110       0.151
##      Black                 253 (20.5)     71 (21.5)
##      Hispanic              220 (17.9)     64 (19.4)
##      Other                 169 (13.7)     59 (17.9)
##      White                 590 (47.9)    136 (41.2)
##   education (%)                                       0.005       0.186
```

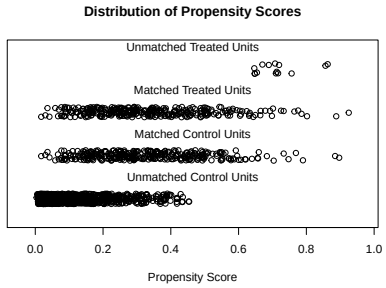
```
##      College                649 (52.7)    157 (47.6)
##      High.School            518 (42.0)    140 (42.4)
##      School                  65 ( 5.3)     33 (10.0)
##   married (%)                                        <0.001       0.282
##      Married                727 (59.0)    194 (58.8)
##      Never.married          201 (16.3)     27 ( 8.2)
##      Previously.married     304 (24.7)    109 (33.0)
##   bmi (mean (SD))        29.14 (7.03)  33.01 (7.65)  <0.001       0.526
```

6.2 Vizualization for Overlap

```
boxplot(PS ~ diabetes, data = analytic,
        lwd = 2, ylab = 'PS')
stripchart(PS ~ diabetes, vertical = TRUE,
           data = analytic, method = "jitter",
           add = TRUE, pch = 20, col = 'blue')
```




```
plot(match.obj, type = "jitter")
```

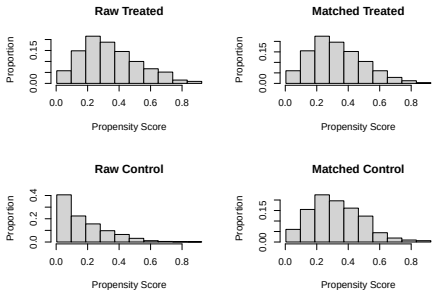



```
## [1] "To identify the units, use first mouse button; to stop, use second."
```

```
## integer(0)
```

Vizualization for assessing overlap issues

```
plot(match.obj, type = "hist")
```



Assessment of Balance: Better than regression diagnostics!

```
matched.data <- match.data(match.obj)
tab1m <- CreateTableOne(vars = baselinevars,
                        strata = "diabetes",
                        data = matched.data,
                        includeNA = TRUE,
                        test = TRUE, smd = TRUE)
```

Compare the similarity of baseline characteristics between treated and untreated subjects in a the propensity score-matched sample.

- In this case, we will compare SMD < 0.1 or not.
- In some literature, other generous values (0.25) are proposed. (Austin, 2011a)

```
print(tab1m, showAllLevels = FALSE, smd = TRUE, test = FALSE)
```

```
##                          Stratified by diabetes
##                           0             1             SMD
##   n                         316           316
##   gender = Male (%)         218 (69.0)    212 (67.1)   0.041
##   age (mean (SD))         63.03 (13.48) 62.67 (12.87)  0.027
##   race (%)                                             0.105
##      Black                   79 (25.0)     68 (21.5)
##      Hispanic                58 (18.4)     61 (19.3)
##      Other                   44 (13.9)     53 (16.8)
##      White                  135 (42.7)    134 (42.4)
##   education (%)                                        0.007
##      College                 153 (48.4)    152 (48.1)
##      High.School             133 (42.1)    134 (42.4)
##      School                   30 ( 9.5)     30 ( 9.5)
##   married (%)                                          0.099
##      Married                 183 (57.9)    186 (58.9)
##      Never.married            20 ( 6.3)     27 ( 8.5)
##      Previously.married      113 (35.8)    103 (32.6)
##   bmi (mean (SD))         32.38 (7.62)  32.63 (7.20)   0.035
```

6.3 SMD vs. P-values

Possible to get p-values to check balance: but strongly discouraged

- P-value based balance assessment can be influenced by sample size

```
print(tab1m, showAllLevels = FALSE, smd = FALSE, test = TRUE)
```

```
##                         Stratified by diabetes
##                          0             1             p      test
##   n                        316           316
##   gender = Male (%)        218 (69.0)    212 (67.1)   0.670
##   age (mean (SD))        63.03 (13.48) 62.67 (12.87)  0.733
##   race (%)                                            0.629
##      Black                  79 (25.0)     68 (21.5)
##      Hispanic               58 (18.4)     61 (19.3)
##      Other                  44 (13.9)     53 (16.8)
##      White                 135 (42.7)    134 (42.4)
##   education (%)                                       0.996
##      College               153 (48.4)    152 (48.1)
##      High.School           133 (42.1)    134 (42.4)
##      School                 30 ( 9.5)     30 ( 9.5)
##   married (%)                                         0.465
##      Married               183 (57.9)    186 (58.9)
##      Never.married          20 ( 6.3)     27 ( 8.5)
##      Previously.married    113 (35.8)    103 (32.6)
##   bmi (mean (SD))        32.38 (7.62)  32.63 (7.20)   0.662
```

Assessment of balance in the matched data

```
smd.res <- ExtractSmd(tab1m)
t(round(smd.res,2))
```

```
##        gender  age race education married  bmi
## 1 vs 2   0.04 0.03 0.11      0.01     0.1 0.03
```

6.4 Variance ratio

- Variance ratios ~ 1 means:
- equal variances in groups
- group balance
- could vary from 1/2 to 2
- other cut-points are suggested as well (0.8 to 1.2)

See Stuart (2010) and Austin (2009)

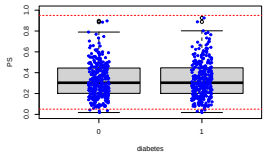
```
require(cobalt)
baltab.res <- bal.tab(x = match.obj, data = analytic,
                      treat = analytic$diabetes,
```

```
                      disp.v.ratio = TRUE)
baltab.res
```

```
## Call
##  matchit(formula = ps.formula, data = analytic, method = "nearest",
##     distance = "logit", replace = FALSE, caliper = 0.2 * sd(logitPS),
##     ratio = 1)
##
## Balance Measures
##                                Type Diff.Adj V.Ratio.Adj
## distance                   Distance   0.0276      1.0992
## gender_Male                  Binary  -0.0190
## age                         Contin.  -0.0278      0.9114
## race_Black                   Binary  -0.0348
## race_Hispanic                Binary   0.0095
## race_Other                   Binary   0.0285
## race_White                   Binary  -0.0032
## education_College            Binary  -0.0032
## education_High.School        Binary   0.0032
## education_School             Binary   0.0000
## married_Married              Binary   0.0095
## married_Never.married        Binary   0.0222
## married_Previously.married   Binary  -0.0316
## bmi                         Contin.   0.0338      0.8928
##
## Sample sizes
##           Control Treated
## All          1232     330
## Matched       316     316
## Unmatched     916      14
```

6.5 Close inspection of boundaries

```
boxplot(PS ~ diabetes, data = matched.data,
        lwd = 2, ylab = 'PS', ylim=c(0,1))
stripchart(PS ~ diabetes, vertical = TRUE,
           data = matched.data, method = "jitter",
           add = TRUE, pch = 20, col = 'blue')
abline(h=c(0+0.05,1-0.05), col = "red", lty = 2)
```



- Sensitivity analysis should be done with trimming.
- Have consequences in interpretation
  - target populstion may be unclear

6.6 Unsatirfactory balance

- Best strategy is to go back to step 2, and make changes in the PS model specification

Chapter 7

Step 4: Outcome modelling

- Some flexibility in choosing outcome model
  - considered independent of exposure modelling
  - some propose double robust approach
  - adjusting imbalanced covariates only?
    - double-adjustment may address residual confounding (Nguyen et al., 2017)

7.1 Crude outcome model

Estimate the effect of treatment on outcomes using propensity score-matched sample

```
fit3 <- glm(cholesterol~diabetes,
            family=binomial, data = matched.data)
publish(fit3)
```

```
##  Variable Units OddsRatio       CI.95  p-value
##  diabetes            0.90 [0.54;1.50]   0.6984
```

7.2 Double-adjustment

Estimate the effect of treatment on outcomes using propensity score-matched sample, and adjust for imbalanced covariate

```
fit3r <- glm(cholesterol~diabetes + race,
            family=binomial, data = matched.data)
publish(fit3r)
```

```
##  Variable    Units OddsRatio       CI.95  p-value
##  diabetes              0.89 [0.54;1.49]   0.6657
```

```
##      race    Black       Ref
##           Hispanic      0.96 [0.46;2.02]   0.9165
##              Other      1.32 [0.63;2.78]   0.4581
##              White      0.58 [0.30;1.13]   0.1095
```

7.3 Adjusted outcome model

Adjust for all covariates, again! (suggested)

```
out.formula <- as.formula(paste("cholesterol", "~ diabetes +",
                                paste(baselinevars,
                                      collapse = "+")))
out.formula
```

```
## cholesterol ~ diabetes + gender + age + race + education + married +
##     bmi
```

```
fit3b <- glm(out.formula,
           family=binomial, data = matched.data)
publish(fit3b)
```

```
##   Variable              Units OddsRatio       CI.95     p-value
##   diabetes                           0.86 [0.51;1.46]   0.5794126
##     gender             Female       Ref
##                          Male      0.38 [0.21;0.69]   0.0012767
##        age                           0.95 [0.93;0.97]     < 1e-04
##       race              Black       Ref
##                        Hispanic      0.72 [0.31;1.65]   0.4346787
##                           Other      0.77 [0.34;1.73]   0.5224157
##                           White      0.51 [0.25;1.04]   0.0649791
##  education            College       Ref
##                     High.School      0.70 [0.39;1.24]   0.2215142
##                          School      0.93 [0.35;2.43]   0.8791455
##    married            Married       Ref
##                  Never.married      0.48 [0.15;1.54]   0.2173180
##             Previously.married      0.84 [0.45;1.57]   0.5900732
##        bmi                           0.93 [0.89;0.97]   0.0005547
```

The above analysis do not take matched pair into consideration while regressing.

7.4 Other cosiderations for outcome model

Literature proposes different strategies:

- do not control for pairs / clusters
  - use `glm` as is
- control for pairs / clusters

  - use `cluster` option (preferred)
  - use GEE or
  - use conditional logistic

Here is an example using cluster option:

```
require(jtools)
summ(fit3b, rubust = "HC0", confint = TRUE, digists = 3,
     cluster = "subclass", model.info = FALSE,
     model.fit = FALSE, exp = TRUE)
```

	exp(Est.)	2.5%	97.5%	z val.	p
(Intercept)	100.02	8.74	1144.55	3.70	0.00
diabetes	0.86	0.51	1.46	-0.55	0.58
genderMale	0.38	0.21	0.69	-3.22	0.00
age	0.95	0.93	0.97	-4.47	0.00
raceHispanic	0.72	0.31	1.65	-0.78	0.43
raceOther	0.77	0.34	1.73	-0.64	0.52
raceWhite	0.51	0.25	1.04	-1.85	0.06
educationHigh.School	0.70	0.39	1.24	-1.22	0.22
educationSchool	0.93	0.35	2.43	-0.15	0.88
marriedNever.married	0.48	0.15	1.54	-1.23	0.22
marriedPreviously.married	0.84	0.45	1.57	-0.54	0.59
bmi	0.93	0.89	0.97	-3.45	0.00

Standard errors: MLE

- Bootstrap for matched pairfor WOR (Austin and Small, 2014)
  - may not be appropriate for WR

7.5 Estimate obtained

- The example compared `diabetic` (a treated group; target) vs `Not diabetic` (untreated).
- Thc corresponding treatment effect estimate is known as
  - Average Treatment Effects on the Treated (**ATT**)
- Other estimates from PS analysis (e.g., PS weighting) are possible that compared the whole population
  - what if everyone treated vs. what if nobody was treated (ATE)

Chapter 8

PS vs. Regression

8.1 Data Simulation

Simplified simulation example, so that we know the true parameter θ.

Y : Outcome	Cholesterol levels (continuous)
A : Exposure	Diabetes
L : Known Confounders	age (continuous)

- Confounder L (continuous)
  - Logit $L \sim$ N(mean $= 10$, sd $= 1$)
- Treatment A (binary 0/1)
  - Logit $P(A = 1) \sim 0.4$ L
- Outcome Y (continuous)
  - Y $\sim$ N(mean $= 3$ L $+ \theta$ A, sd $= 1$)

$\theta = 0.7$

```
require(simcausal)
D <- DAG.empty()
D <- D +
  node("L", distr = "rnorm", mean = 10, sd = 1) +
  node("A", distr = "rbern", prob = plogis(0.4*L)) +
  node("Y", distr = "rnorm", mean = 3 * L + 0.7 * A, sd = 1)
Dset <- set.DAG(D)
```

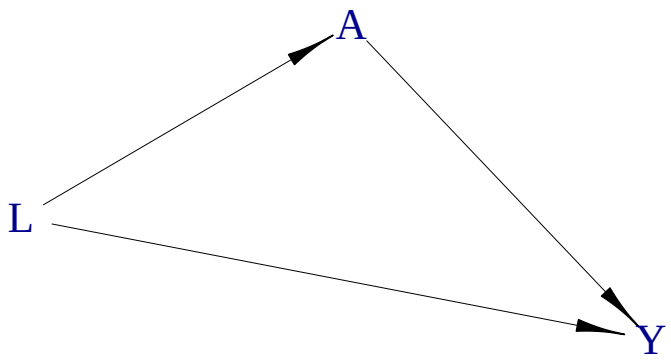
```
plotDAG(Dset, xjitter = 0.1, yjitter = .9,
        edge_attrs = list(width = 0.5, arrow.width = 0.4, arrow.size = 1.7),
        vertex_attrs = list(size = 18, label.cex = 1.8))
```

```
## using the following vertex attributes:
```

```
## 181.8NAdarkbluenone0

## using the following edge attributes:

## 0.50.41.7black1
```




```
# Data generating function
fnc <- function(n = 10, seedx = 123){
  require(simcausal)
  set.seed(seedx)
  D <- DAG.empty()
  D <- D +
    node("L", distr = "rnorm", mean = 10, sd = 1) +
    node("A", distr = "rbern", prob = plogis(0.4*L)) +
    node("Y", distr = "rnorm", mean = 3 * L + 0.7 * A, sd = 1)
  Dset <- set.DAG(D)
  A1 <- node("A", distr = "rbern", prob = 1)
  Dset <- Dset + action("A1", nodes = A1)
  A0 <- node("A", distr = "rbern", prob = 0)
  Dset <- Dset + action("A0", nodes = A0)
  Cdat <- sim(DAG = Dset, actions = c("A1", "A0"), n = n, rndseed = 123)
  generated.data <- round(cbind(Cdat$A1[c("ID", "L", "Y")],Cdat$A0[c("Y")]),2)
  names(generated.data) <- c("ID", "L", "Y1", "Y0")
  generated.data <- generated.data[order(generated.data$L, generated.data$ID),]
  generated.data$A <- sample(c(0,1),n, replace = TRUE)
  generated.data$Y <- ifelse(generated.data$A==0, generated.data$Y0, generated.data$Y1)
```

```
  counterfactual.dataset<- generated.data[order(generated.data$ID) , ][c("ID","L","A","Y1","Y0")]
  observed.dataset<- generated.data[order(generated.data$ID) , ][c("ID","L","A","Y")]
  return(list(counterfactual=counterfactual.dataset,
              observed=observed.dataset))
}
```

10 observations from the data generation:

```
result.data <- fnc(n=10)
```

```
result.data
```

```
## $counterfactual
##    ID     L A    Y1    Y0
## 1   1  9.44 0 30.24 29.54
## 2   2  9.77 1 30.37 29.67
## 3   3 11.56 1 35.78 35.08
## 4   4 10.07 0 31.02 30.32
## 5   5 10.13 0 30.53 29.83
## 6   6 11.72 0 37.63 36.93
## 7   7 10.46 1 32.58 31.88
## 8   8  8.73 1 24.94 24.24
## 9   9  9.31 1 29.34 28.64
## 10 10  9.55 1 28.89 28.19
##
## $observed
##    ID     L A     Y
## 1   1  9.44 0 29.54
## 2   2  9.77 1 30.37
## 3   3 11.56 1 35.78
## 4   4 10.07 0 30.32
## 5   5 10.13 0 29.83
## 6   6 11.72 0 36.93
## 7   7 10.46 1 32.58
## 8   8  8.73 1 24.94
## 9   9  9.31 1 29.34
## 10 10  9.55 1 28.89
```

8.2 Treatment effect from counterfactuals

True θ can be obtained from counterfactual data:

```
result.data$counterfactual$TE <- result.data$counterfactual$Y1- result.data$counterfactual$Y0
result.data$counterfactual
```

```
##    ID     L A    Y1    Y0  TE
## 1   1  9.44 0 30.24 29.54 0.7
```

```
## 2   2  9.77 1 30.37 29.67 0.7
## 3   3 11.56 1 35.78 35.08 0.7
## 4   4 10.07 0 31.02 30.32 0.7
## 5   5 10.13 0 30.53 29.83 0.7
## 6   6 11.72 0 37.63 36.93 0.7
## 7   7 10.46 1 32.58 31.88 0.7
## 8   8  8.73 1 24.94 24.24 0.7
## 9   9  9.31 1 29.34 28.64 0.7
## 10 10  9.55 1 28.89 28.19 0.7
```

8.3 Treatment effect from Regression

What happens in observed data for a sample of size 10?

```
round(coef(glm(Y ~ A, family="gaussian", data=result.data$observed)),2)
```

```
## (Intercept)           A
##       31.65       -1.34
```

```
round(coef(glm(Y ~ A + L, family="gaussian", data=result.data$observed)),2)
```

```
## (Intercept)           A           L
##       -5.17        0.24        3.56
```

What happens in observed data for a sample of size 10000?

```
result.data <- fnc(n=10000)
```

```
round(coef(glm(Y ~ A, family="gaussian", data=result.data$observed)),2)
```

```
## (Intercept)           A
##       29.98        0.70
```

```
round(coef(glm(Y ~ A + L, family="gaussian", data=result.data$observed)),2)
```

```
## (Intercept)           A           L
##       -0.07        0.70        3.01
```

8.4 Treatment effect from PS

Propensity score model fitting:

```
require(MatchIt)
match.obj <- matchit(A ~ L, method = "nearest",
                     data = result.data$observed,
                     distance = 'logit',
                     caliper = 0.001,
                     replace = FALSE,
```

```
                     ratio = 1)
match.obj
```

```
## A matchit object
##  - method: 1:1 nearest neighbor matching without replacement
##  - distance: Propensity score [caliper]
##              - estimated with logistic regression
##  - caliper: <distance> (0)
##  - number of obs.: 10000 (original), 8306 (matched)
##  - target estimand: ATT
##  - covariates: L
```

Results from step 4: crude

```
matched.data <- match.data(match.obj)
```

Results from step 4: adjusted

```
round(coef(glm(Y ~ A, family="gaussian", data=matched.data)),2)
```

```
## (Intercept)           A
##       29.97        0.69
```

```
round(coef(glm(Y ~ A+L, family="gaussian", data=matched.data)),2)
```

```
## (Intercept)           A           L
##       -0.10        0.69        3.01
```

8.5 Non-linear Model

8.5.1 Data generation

Y : Outcome	Cholesterol levels (continuous)
A : Exposure	Diabetes
L : Known Confounders	age (continuous)

- Confounder L (continuous)
  - Logit $L \sim$ N(mean $= 10$, sd $= 1$)
- Treatment A (binary 0/1)
  - Logit $P(A = 1) \sim 0.4$ L
- Outcome Y (continuous)
  - Y $\sim$ N(mean $= 3\ L^3 + \theta$ A, sd $= 1$)

The only difference is L^3 instead of L in the outcome mode.

Again, $\theta = 0.7$

```
# Data generating function
fnc2 <- function(n = 10, seedx = 123){
  require(simcausal)
  set.seed(seedx)
  D <- DAG.empty()
  D <- D +
    node("L", distr = "rnorm", mean = 10, sd = 1) +
    node("A", distr = "rbern", prob = plogis(0.4*L)) +
    node("Y", distr = "rnorm", mean = 3 * L^3 + 0.7 * A, sd = 1)
  Dset <- set.DAG(D)
  A1 <- node("A", distr = "rbern", prob = 1)
  Dset <- Dset + action("A1", nodes = A1)
  A0 <- node("A", distr = "rbern", prob = 0)
  Dset <- Dset + action("A0", nodes = A0)
  Cdat <- sim(DAG = Dset, actions = c("A1", "A0"), n = n, rndseed = 123)
  generated.data <- round(cbind(Cdat$A1[c("ID", "L", "Y")],Cdat$A0[c("Y")]),2)
  names(generated.data) <- c("ID", "L", "Y1", "Y0")
  generated.data <- generated.data[order(generated.data$L, generated.data$ID),]
  generated.data$A <- sample(c(0,1),n, replace = TRUE)
  generated.data$Y <- ifelse(generated.data$A==0, generated.data$Y0, generated.data$Y1)
  counterfactual.dataset<- generated.data[order(generated.data$ID) , ][c("ID","L","A","
  observed.dataset<- generated.data[order(generated.data$ID) , ][c("ID","L","A","Y")]
  return(list(counterfactual=counterfactual.dataset,
              observed=observed.dataset))
}
```

8.5.2 Regression

```
result.data <- fnc2(n=10000)
```

Crude estimates

```
round(coef(glm(Y ~ A, family="gaussian", data=result.data$observed)),2)
```

```
## (Intercept)           A
##     3094.49      -13.16
```

Adjusted estimates

```
fit <- glm(Y ~ A + L, family="gaussian", data=result.data$observed)
round(coef(fit),2)
```

```
## (Intercept)           A           L
##    -6002.42       -1.25      909.32
```

- In regression adjustments, the results could be subject to "model extrapolation" based on linearity assumption.

- It is sometimes difficult to know whether the adjusted effect is based on extrapolation.
- Especially true in observational settings.
- PS may not need such linearity assumption (when non-parametric approaches used for prediction).
  - Don't necessarily mean non-parametric approaches are the best option though!

8.5.3 PS

Matching with PS

```
match.obj <- matchit(A ~ L, method = "nearest",
                     data = result.data$observed,
                     distance = 'logit',
                     replace = FALSE,
                     caliper = 0.001,
                     ratio = 1)
match.obj
```

```
## A matchit object
##  - method: 1:1 nearest neighbor matching without replacement
##  - distance: Propensity score [caliper]
##              - estimated with logistic regression
##  - caliper: <distance> (0)
##  - number of obs.: 10000 (original), 8282 (matched)
##  - target estimand: ATT
##  - covariates: L
```

```
matched.data <- match.data(match.obj)
```

Results from step 4: crude

```
round(coef(glm(Y ~ A, family="gaussian", data=matched.data)),2)
```

```
## (Intercept)           A
##     3070.64        0.72
```

Results from step 4: adjusted

```
round(coef(glm(Y ~ A+L, family="gaussian", data=matched.data)),2)
```

```
## (Intercept)           A           L
##    -5980.08        0.72      905.62
```

8.5.4 Machine learning

Using gradient boosted method for PS estimation

```
require(twang)
result.data$observed$S <- 0
ps.gbm <- ps(A ~ L + S,data = result.data$observed,estimand = "ATT",n.trees=1000)
names(ps.gbm)
summary(ps.gbm$ps$es.mean.ATT)
result.data$observed$ps <- ps.gbm$ps$es.mean.ATT
```

Matching with PS generated from gradient boosted method

```
require(Matching)
match.obj2 <- Match(Y=result.data$observed$Y, Tr=result.data$observed$A,
                    X=result.data$observed$ps, M=1, caliper = 0.001,
                    replace=FALSE)
summary(match.obj2)
```

```
##
## Estimate...  1.5255
## SE.........  1.6227
## T-stat.....  0.94006
## p.val......  0.34719
##
## Original number of observations..............  10000
## Original number of treated obs...............  4968
## Matched number of observations...............  4520
## Matched number of observations  (unweighted).  4520
##
## Caliper (SDs)........................................   0.001
## Number of obs dropped by 'exact' or 'caliper'  448
```

```
matched.data2 <- result.data$observed[c(match.obj2$index.treated, match.obj2$index.con
mb <- MatchBalance(A~L, data=result.data$observed, match.out=match.obj2, nboots=10)
```

Results from step 4: crude

```
round(coef(glm(Y ~ A, family="gaussian", data=matched.data2)),2)
```

```
## (Intercept)           A
##     3089.17        1.53
```

Results from step 4: adjusted

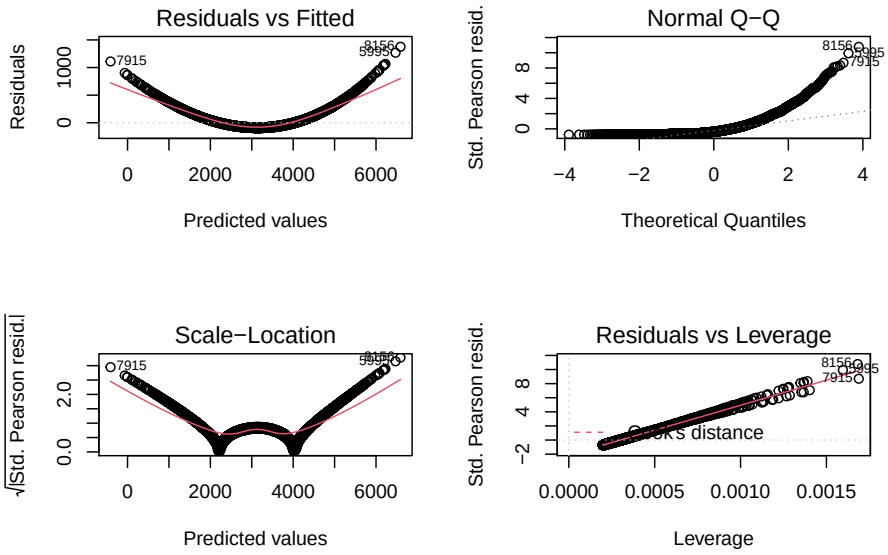
```
round(coef(glm(Y ~ A+L, family="gaussian", data=matched.data2)),2)
```

```
## (Intercept)           A           L
##    -6025.34        1.55      910.70
```

8.5.5 Regression is doomed?

Not really. Always a god idea to check the diagnostic plots to find any indication of assumption violation:

```
par(mfrow=c(2,2))
plot(fit)
```



Residual plot has a pattern!

	Powerful machine learning method is good at prediction.
	Propensity score methods rely on obtaining good balance.
	Always a good idea to check analysis with multiple sensitivity analysis.

8.6 Common misconception

- PS results = 'causal';
- regression = 'non-causal'.

No. 'Results from both methods should lead to the same conclusions.' (D'Agostino Jr, 1998)

When the results deviate, important to investigate why!

8.7 Benifits of PS

- **Intuitive**: compare two similar groups

- **2-step process**
  - Encourages researchers to think about the treatment generation process
  - Fit outcome model with only important variables.
  - Allowing to think more about design stage (nice separation from outcome model building process).
- Fit **rich PS model** (with higher order terms); focusing on prediction; worry less about overparameterization.
  - Non-parametric (ML) approaches can be used to relax linearity assumption in estimating PS.
  - See more on Lee et al. (2010), Pirracchio et al. (2015), Alam et al. (2019)
- **Reduce dimension**, helpful when exposure frequent but outcome rare (event per variable).
  - Smaller outcome model may be helpful in diagnostic checks.
- **Diagnostics**
  - Diagnostics (balance checking) much easier compared to residual plot/influence
  - Graphical comparison helps identify areas of non-overlap.

8.8 Limitations of PS

- Matching population vs. target population: often not the same.
  - PS matching may give effect estimate of a subset, which may be difficult to identify in the actual population!
- PS can do nothing about unmeasured confounding, neither can outcome regression.

Chapter 9

Reporting Guidelines

While writing journal articles or reports, what are the components we should report?

9.1 Discipline-specific Reviews

- Propensity score matching most popular
  - Cardiovascular (Austin, 2007),
  - Infective endocarditis,
  - Intensive care
  - Critical care,
  - anesthesiology,
  - Sepsis,
  - Psychology
  - Cancer (Yao et al., 2017),
  - Multiple sclerosis (Karim et al., 2020)

9.2 Suggested Guidelines

Population	Be specific about population of interest - ATT vs. ATE - exclusion criteria
Intervention	Be specific about exposure - no multiple version of treatment - no interference - comparator
Covariates	How variables are selected - Any important variables not measured? Proxy? - Large list of covariates? See King and Nielsen (2019)

PS Model	Model selection - interaction or polynomials - logistic vs. machine learning - Residual imbalance and refit PS model
PS approach	Why PS matching (or other approach) was selected?
Sample size	Reduction % of the matched data: major issue!
Diagnostics	Overlap vs. balance assessments - numeric and visual
Sensitivity analysis	- unmeasured confounder / hdPS - any positivity issue? Deleting extremes has consequences! - ad-hoc methods: truncation / trimming: bias-variance trade-off
Subgroup analysis	Refit within each group for matching - See Ali et al. (2019) for a more complete list
Missing data	Report clearly about missing data - how missing data handled
Software	Report software

9.3 Software

- Useful R packages
  - MatchIt
  - cobalt
  - Matching
  - twang
- Also see
  - Elizabeth Stuart's Propensity Score Software Page for SAS, STATA, SPSS, Excel packages

9.4 Further Resources

- My workshop page
- My YouTube channel for related PS materials
- Teaching by WebApps: particularly this one.

References

Bibliography

Alam, S., Moodie, E. E., and Stephens, D. A. (2019). Should a propensity score model be super? the utility of ensemble procedures for causal adjustment. *Statistics in medicine*, 38(9):1690–1702.

Ali, M. S., Prieto-Alhambra, D., Lopes, L. C., Ramos, D., Bispo, N., Ichihara, M. Y., Pescarini, J. M., Williamson, E., Fiaccone, R. L., Barreto, M. L., et al. (2019). Propensity score methods in health technology assessment: principles, extended applications, and recent advances. *Frontiers in pharmacology*, 10:973.

Austin, P. C. (2007). Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. *The Journal of thoracic and cardiovascular surgery*, 134(5):1128–1135.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25):3083–3107.

Austin, P. C. (2011a). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3):399–424.

Austin, P. C. (2011b). A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate behavioral research*, 46(1):119–151.

Austin, P. C. and Small, D. S. (2014). The use of bootstrapping when using propensity-score matching without replacement: a simulation study. *Statistics in medicine*, 33(24):4306–4319.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., and Stürmer, T. (2006). Variable selection for propensity score models. *American journal of epidemiology*, 163(12):1149–1156.

D'Agostino Jr, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in medicine*, 17(19):2265–2281.

Gautret, P., Lagier, J.-C., Parola, P., Meddeb, L., Mailhe, M., Doudier, B., Courjon, J., Giordanengo, V., Vieira, V. E., Dupont, H. T., et al. (2020). Hydroxychloroquine and azithromycin as a treatment of covid-19: results of an open-label non-randomized clinical trial. *International journal of antimicrobial agents*, 56(1):105949.

Karim, M. E., Pang, M., and Platt, R. W. (2018). Can we train machine learning methods to outperform the high-dimensional propensity score algorithm? *Epidemiology*, 29(2):191–198.

Karim, M. E., Pellegrini, F., Platt, R. W., Simoneau, G., Rouette, J., and de Moor, C. (2020). The use and quality of reporting of propensity score methods in multiple sclerosis literature: A review. *Multiple Sclerosis Journal*, page 1352458520972557.

King, G. and Nielsen, R. A. (2019). Why propensity scores should not be used for matching.

Lee, B. K., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in medicine*, 29(3):337–346.

Nguyen, T.-L., Collins, G. S., Spence, J., Daurès, J.-P., Devereaux, P., Landais, P., and Le Manach, Y. (2017). Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. *BMC medical research methodology*, 17(1):1–8.

Pirracchio, R., Petersen, M. L., and Van Der Laan, M. (2015). Improving propensity score estimators' robustness to model misspecification using super learner. *American journal of epidemiology*, 181(2):108–119.

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

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.

Yao, X. I., Wang, X., Speicher, P. J., Hwang, E. S., Cheng, P., Harpole, D. H., Berry, M. F., Schrag, D., and Pang, H. H. (2017). Reporting and guidelines in propensity score analysis: a systematic review of cancer and cancer surgical studies. *JNCI: Journal of the National Cancer Institute*, 109(8):djw323.