STA305/1004 - Class 10

October 15, 2019

Today's class

- Estimating the propensity score

Today's class

- Estimating the propensity score
- The balancing property of the propensity score

Today's class

- Estimating the propensity score
- The balancing property of the propensity score
- Assessing balance

Today's class

- Estimating the propensity score
- The balancing property of the propensity score
- Assessing balance
- Ignorable treatment assignment and the propensity score

Today's class

- Estimating the propensity score
- The balancing property of the propensity score
- Assessing balance
- Ignorable treatment assignment and the propensity score
- Three methods that use the propensity score to reduce bias: matching; stratification; and regression adjustment

The propensity score

- Covariates are pre-treatment variables and take the same value for each unit no matter which treatment is applied.

The propensity score is

$$e(\mathbf{x}) = P\left(T = 1|\mathbf{x}\right),$$

where $\mathbf{x}$ are observed covariates.

The i^{th} propensity score is the probability that a unit receives treatment given all the information, recorded as covariates, that is observed before the treatment.

The propensity score

- Covariates are pre-treatment variables and take the same value for each unit no matter which treatment is applied.
- For example, pre-treatment blood pressure or pre-test reading level are not influenced by a treatment that would alter blood pressure or reading level.

The propensity score is

$$e(\mathbf{x}) = P\left(T = 1 | \mathbf{x}\right),$$

where $\mathbf{x}$ are observed covariates.

The i^{th} propensity score is the probability that a unit receives treatment given all the information, recorded as covariates, that is observed before the treatment.

The propensity score

In experiments the propensity scores are known. In observational studies they can be estimated using models such as logistic regression where the outcome is the treatment indicator and the predictors are all the confounding covariates.

The propensity score

- Consider a study that plans to use a doctor's medical records to compare two treatments ($T = 0$ and $T = 1$) given for a certain condition.

The propensity score

- Consider a study that plans to use a doctor's medical records to compare two treatments ($T = 0$ and $T = 1$) given for a certain condition.
- Treatments were not assigned to patients randomly, but were based on various measured and unmeasured patient factors.

Logistic Regression

- The logistic regression model with one covariate x is:

$$log\left(P(T_i = 1)/P(T_i = 0)\right) = \beta_0 + \beta_1 x_i$$

Logistic Regression

- The logistic regression model with one covariate x is:

$$log\left(P(T_i = 1)/P(T_i = 0)\right) = \beta_0 + \beta_1 x_i$$

See details on next slide

- The logistic regression model with k covariates $x_1, x_2, ..., x_k$ is

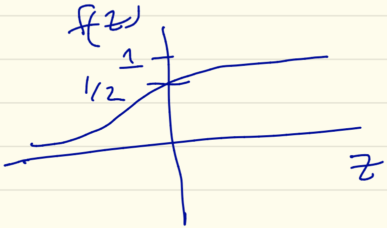
$$log\left(P(T_i = 1)/P(T_i = 0)\right) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_1 x_{ik}$$

Let T be trt. indicator : $T = \begin{cases} 1 & \text{treated} \\ 0 & \text{not treated.} \end{cases}$

Want to model $P(T=1) = \beta_0 + \beta_1 x$

If linear regression is used then predicted values may be outside $[0,1]$. This is an issue $\because$ $P(T=1) \in [0,1]$. $\therefore$ we will not use linear regression.

Instead model : $P(T=1) = \underbrace{\dfrac{1}{1+e^{-z}}}_{\text{logistic function.}}$, $z = \beta_0 + \beta_1 x$



$$1+e^{-z} = \frac{1}{P(T=1)}$$

$$-z = \log\left(\frac{1-P(T=1)}{P(T=1)}\right)$$

$$z = \log\left(\frac{P(T=1)}{1-P(T=1)}\right)$$

$$\beta_0 + \beta_1 x = \log\left(\frac{P(T=1)}{1-P(T=1)}\right)$$

$$= \log\left(P(T=1) / P(T=0)\right)$$

β_0, β_1 in logistic regression are usually estimated using maximum likelihood estimation. **NOT** least squares.

Parameter Estimates from Logistic Regression

$x = \begin{cases} 1 & \text{blue eyes} \\ 0 & \text{not blue eyes.} \end{cases}$

$\log\left(\frac{P(T=1)}{P(T=0)}\right) = \beta_0 + \beta_1 x$

when $x = 1$

(1) $\log\left(\frac{P(T=1)}{P(T=0)}\right) = \beta_0 + \beta_1$

- In a logistic model with one binary covariate the parameter estimate of β_1 is:

odds ratio →

$$\frac{(P(T=1|x=1)/P(T=0|x=1))}{(P(T=1|x=0)/P(T=0|x=0))} = exp(\beta_1)$$

when $x = 0$

(2) $\log\left(\frac{P(T=1)}{P(T=0)}\right) = \beta_0$

(1) − (2)

$= \log\left(\frac{P(T=1|x=1)}{P(T=0|x=1)}\right) - \log\left(\frac{P(T=1|x=0)}{P(T=0|x=0)}\right) = \beta_1$

$\exp(\beta_1) = \frac{P(T=1|x=1)}{P(T=0|x=1)} \Big/ \frac{P(T=1|x=0)}{P(T=0|x=0)}$

(A) $$\frac{P(T=1 \mid x=1)}{P(T=0 \mid x=1)}$$ = odds of recieving treatment when $x=1$.

$$\text{odds} = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1-P(A)}$$

(B) $$\frac{P(T=1 \mid x=0)}{P(T=0 \mid x=0)}$$ = odds of recieving treatment when $x=0$

$\therefore$ (A) / (B) is the odds ratio.

Parameter Estimates from Logistic Regression

- In a logistic model with one binary covariate the parameter estimate of β_1 is:

$$\frac{(P(T=1|x=1)/P(T=0|x=1))}{(P(T=1|x=0)/P(T=0|x=0))} = exp(\beta_1)$$

- $exp(\beta_1)$ is the odds ratio comparing those with $x = 1$ to those with $x = 0$.

Predicted probabilities from Logistic Regression

$\log\left(\frac{p}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 x$, $\hat{\beta}_0, \hat{\beta}_1$ are estimates of β_0, β_1

$p = P(Y=1)$

Solve for p

$\frac{p}{1-p} = e^{\hat{\beta}_0 + \beta_1 x}$

- In a logistic model with one binary covariate the predicted probabilities can be calculated using the fitted model:

$$\hat{p}_i = \frac{exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1}\right)}{1 + exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1}\right)}$$

$\frac{p}{1-p} = z$

$p = z - zp$

$p + zp = z$

$p(1+z) = z$

$p = \frac{z}{1+z}$

The propensity score

- The patient factors that were measured are age (x_1), sex (x_2), and health status before treatment (x_3).

$$log\left(\frac{p_i}{1-p_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3},$$

where $p_i = P(T_i = 1)$.

The propensity score

- The patient factors that were measured are age (x_1), sex (x_2), and health status before treatment (x_3).
- The propensity score can be estimated for each patient by fitting a logistic regression model with treatment as the dependent variable and x_1, x_2, x_3 as the predictor variables.

$$log\left(\frac{p_i}{1-p_i}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3},$$

where $p_i = P(T_i = 1)$.

plug in x_{ik}, $k=1,2,3$ to obtain predicted probabilities.

The propensity score

- The predicted probabilities from the above equation are estimates of the propensity score for each patient.

$$\hat{p}_i = \frac{exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3}\right)}{1 + exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{\beta}_3 x_{i3}\right)}$$

The propensity score in Smoking Cessation Study

The propensity score for each subject in smoking and weight gain study can be estimated by fitting a logistic regression model.

The propensity score in Smoking Cessation Study - The Data

Stop Smoking is Yes

Stop Smoking is No

	Cessation (T=1)	No cessation (T=0)
age, years	46.2	42.8
men, %	54.6	46.6
white, %	91.1	85.4
university, %	15.4	9.9
weight, kg	72.4	70.3
Cigarettes/day	18.6	21.2
year smoking	26.0	24.1
little/no exercise, %	40.7	37.9
inactive daily life, %	11.2	8.9

Predict Smoking Cessation - Propensity Score Model

$Y \sim x_1 + x_2 + x_3 + \cdots$

R function to estimate logistic regression model.

```
prop.model <- glm(qsmk ~ as.factor(sex) + as.factor(race) +
                  age + as.factor(education.code) + smokeintensity +
                  smokeyrs  + as.factor(exercise) + as.factor(active) +
                  wt71, family = binomial(), data = nhefshwdat)

#Summary of propensity score model
summary(prop.model)
```

this specifies that the glm(.) should perform logistic regression.

NB: `qsmk = 1` is smoking cessation and qsmk=0 is not smoking cessation.

Predict Smoking Cessation - Propensity Score Model

```
                               Estimate  Std. Error    z value
(Intercept)                -2.401228039 0.484016356 -4.9610473
as.factor(sex)1            -0.499080121 0.146530691 -3.4059767
as.factor(race)1           -0.778222994 0.207031619 -3.7589572
age                         0.046207220 0.009889326  4.6724338
as.factor(education.code)2 -0.065716379 0.196122828 -0.3350777
as.factor(education.code)3  0.052634524 0.175523000  0.2998725
as.factor(education.code)4  0.108653058 0.269190883  0.4036283
as.factor(education.code)5  0.466164550 0.224105901  2.0801083
smokeintensity             -0.026527450 0.005664293 -4.6832762
smokeyrs                   -0.028491730 0.010008629 -2.8467165
as.factor(exercise)1        0.359556747 0.178603430  2.0131570
as.factor(exercise)2        0.422771538 0.185656969  2.2771649
as.factor(active)1          0.044927909 0.131555137  0.3415139
as.factor(active)2          0.158150602 0.213435405  0.7409764
wt71                        0.006099273 0.004368231  1.3962800
                               Pr(>|z|)
(Intercept)                7.011411e-07
as.factor(sex)1            6.592780e-04
as.factor(race)1           1.706230e-04
age                        2.976515e-06
as.factor(education.code)2 7.375665e-01
as.factor(education.code)3 7.642744e-01
```

$H_0: \beta_i = 0$

$H_a: \beta_i \neq 0$

p-value for testing $H_0: \beta_i = 0$

How do we build a propensity score model?

- Usual tool is logistic regression model for the treatment allocation decision – We therefore want to consider including any variables that have a relationship to the treatment decision (i.e. precede it in time, and are relevant) – No information is included on the actual treatment received, or on the outcome(s).

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony. ✓

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients
6. Thou shalt perform bootstrap analyses to assess shrinkage.

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients
6. Thou shalt perform bootstrap analyses to assess shrinkage.
7. Thou shalt perform regression diagnostics and examine residuals with care.

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients
6. Thou shalt perform bootstrap analyses to assess shrinkage.
7. Thou shalt perform regression diagnostics and examine residuals with care.
8. Thou shalt hold out a sample of thy data for cross-validation.

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients
6. Thou shalt perform bootstrap analyses to assess shrinkage.
7. Thou shalt perform regression diagnostics and examine residuals with care.
8. Thou shalt hold out a sample of thy data for cross-validation.
9. Thou shalt perform external validation on a new sample of data.

Ten commandments of Propensity Model Development

1. Thou shalt value parsimony.
2. Thou shalt examine thy predictors for collinearity.
3. Thou shalt test all thy predictors for statistical significance.
4. Thou shalt have ten times as many predictors as subjects.
5. Thou shalt examine thy regression coefficients
6. Thou shalt perform bootstrap analyses to assess shrinkage.
7. Thou shalt perform regression diagnostics and examine residuals with care.
8. Thou shalt hold out a sample of thy data for cross-validation.
9. Thou shalt perform external validation on a new sample of data.
10. Thou shalt ignore commandments 1 through 9 and instead ensure that the model adequately balances covariates.

Don't worry about these issues when building a logistic regression model to estimate propensity scores.

Propensity model development

1. Diagnostics for the successful prediction of probabilities and parameter estimates underlying those probabilities

In propensity score model development the second point is important, but the first is not important .

Propensity model development

1. Diagnostics for the successful prediction of probabilities and parameter estimates underlying those probabilities
2. Diagnostics for the successful design of observational studies based on estimated propensity scores.

In propensity score model development the second point is important, but the first is not important .

Propensity model development

- All covariates that subject matter experts (and subjects) judge important when selecting treatments.

Propensity model development

- All covariates that subject matter experts (and subjects) judge important when selecting treatments.
- All covariates that relate to treatment and outcome, including any covariate that improves prediction (of exposure group).

Propensity model development

- All covariates that subject matter experts (and subjects) judge important when selecting treatments.
- All covariates that relate to treatment and outcome, including any covariate that improves prediction (of exposure group).
- As much "signal" as possible.

Propensity score in smoking cessation study

The propensity score for each subject is $\hat{p}_i$ is the predicted probability of quitting smoking from the logistic regression model. The predicted probabilities are obtained using `predict()`.

```
#Propensity scores for each subject
p.qsmk.obs <- predict(prop.model, type = "response")
p.qsmk.obs[1:4] # print out first four pred probs
```

```
        1         2         3         4 
0.1239035 0.1597305 0.1599358 0.3106921 
```

logistic regression model.

This function uses the data to calculate the predicted prob. of treatment using all the covariates.

Propensity score in smoking cessation study

Truth - observed value of qsmk

qsmk = 1 quit = yes
0 quit = no

Subject	Quit Smoking	Estimated Propensity Score
1	0	0.12
2	0	0.16
3	0	0.16
4	0	0.31
5	0	0.32
6	0	0.17
7	0	0.24
8	0	0.26
9	0	0.30
10	0	0.29
11	1	0.26
12	0	0.19

Subject 1's estimated probability of quitting smoking is 0.12 (so the estimated probability of not quitting smoking is 1- 0.12=0.82) and subject 11's estimated probability of quitting smoking (propensity score) is 0.26 (so the estimated probability of not quitting smoking is 1-.26=0.74).

Propensity score in smoking cessation study

```r
#predicted value for the first subject
p1 <- predict.glm(prop.model)[1]
p1
```

```
##         1
## -1.955973
```

$$\frac{p}{1-p} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots \hat{\beta}_q x_q}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_q x_q}}$$

```r
exp(p1)/(1 + exp(p1))
```

```
##         1
## 0.1239035
```

```r
# use type="response" to get predicted prob
predict.glm(prop.model,type = "response")[1]
```

```
##         1
## 0.1239035
```

The balancing property of the propensity score

The balancing property of the propensity score says that treated ($T = 1$) and control ($T = 0$) subjects with the same propensity score $e(\mathbf{x})$ have the same distribution of the observed covariates, $\mathbf{x}$,

$$P(\mathbf{x} | T = 1, e(\mathbf{x})) = P(\mathbf{x} | T = 0, e(\mathbf{x}))$$

or

independent.

$$T \perp \mathbf{x} | e(\mathbf{x}).$$

This means that treatment is independent of the observed covariates conditional on the propensity score.

The balancing property of the propensity score

The balancing property says that if two units, i and j, are paired, one of whom is treated, $T_i + T_j = 1$, so that they have the same value of the propensity score $e(\mathbf{x}_i) = e(\mathbf{x}_j)$, then they may have different values of the observed covariate,

$$\mathbf{x}_i \neq \mathbf{x}_j,$$

but in this pair the specific value of the observed covariate will be unrelated to the treatment assignment since

Balancing property.

$$P\left(\mathbf{x} | T = 1, e(\mathbf{x})\right) = P\left(\mathbf{x} | T = 0, e(\mathbf{x})\right)$$

Distribution of covariates in T=1 group with propensity score $e(\underset{\sim}{x})$

Distribution of covariates in T=0 group with propensity score $e(\underset{\sim}{x})$

In a randomized study the distr of unobserved cov. will be the same in two groups

The balancing property of the propensity score

Respond at **PollEv.com/nathantaback**
Text **NATHANTABACK** to **37607** once to join, then **A, B, C, or D**

Pick the answer that makes the following statement True. The balancing property of the propensity score implies that

x	Logistic regression can be used to calculate propensity scores.	**A**	21%
X 35%	Observational studies can be turned into randomized studies if they are balanced using the propensity score.	**B**	24%
✓ 66%	The observed covariate distribution is the same in the treated and untreated groups.	**C**	37%
x	If a treated and untreated experimental unit is matched on the propensity score then the two units must have the same age (assuming age was one of the observed covariates).	**D**	18%

45

Total Results: 0

38

The balancing property of the propensity score

The propensity scores for subject's 10 and 18 in the smoking cessation study are

	Quit Smoking	Estimated Propensity Score
10	0	0.2941244
18	1	0.3197956

The difference between the two subject's propensity scores are 0.32-0.29=0.03. This could be set as a "caliper" or "tolerance" for what are considered equal propensity scores.

The covariates for each subject are

	age	sex	race	edu	smkint	smkyrs	exer	active	wt1971	qsmk
10	43	0	0	2	20	25	2	1	62.26	0
18	48	1	0	3	2	30	1	1	62.03	1

The balancing property of the propensity score

- If many pairs are formed this way then the the distribution of the observed covariates will look about the same in the treated and control groups.

How can the degree of balance in the covariate distributions between treated and control units be assessed?

The balancing property of the propensity score

- If many pairs are formed this way then the the distribution of the observed covariates will look about the same in the treated and control groups.
- Individuals in matched pairs will typically have different values of x.

How can the degree of balance in the covariate distributions between treated and control units be assessed?

The balancing property of the propensity score

i.e., balancing property.

- If many pairs are formed this way then the the distribution of the observed covariates will look about the same in the treated and control groups.
- Individuals in matched pairs will typically have different values of x.
- It is difficult to match on 9 covariates at once, it is easy to match on one covariate, the propensity score $e(\mathbf{x})$, and matching on $e(\mathbf{x})$ will tend to balance all 9 covariates.

How can the degree of balance in the covariate distributions between treated and control units be assessed?

The balancing property of the propensity score

If the smoking cessation and smoking groups are balanced using the propensity score then both observed and unobserved covariates will have similar distributions in the two groups. Thus, this observational study has been turned into a randomized study by using propensity score. methods.

Respond at **PollEv.com/nathantaback**

Text **NATHANTABACK** to 37607 once to join, then **A or B**

True	**A**
False	**B**

Total Results: 0

In an observational study it's not possible to guaruntee that the unobserved covariates will have similar distributions in the treated and untreated groups.

Assessing balance

- The difference in average covariate values by treatment status, scaled by their sample standard deviation. This provides a scale-free way to assess the differences.

Assessing balance

- The difference in average covariate values by treatment status, scaled by their sample standard deviation. This provides a scale-free way to assess the differences.
- As a rule-of-thumb, when treatment groups have important covariates that are more than one-quarter or one-half of a standard deviation apart, simple regression methods are unreliable for removing biases associated with differences in covariates (Imbens and Rubin (2015)).

Assessing balance

If $\bar{x}_t, s_t^2$ are the mean and variance of a covariate in the treated group and $\bar{x}_c, s_c^2$ are the mean and variance of a covariate in the control group then the pooled variance is

$$\sqrt{\frac{s_t^2 + s_c^2}{2}}.$$

The absolute pooled standardized difference is,

$$\frac{100 \times |\bar{x}_t - \bar{x}_c|}{\sqrt{\frac{s_t^2+s_c^2}{2}}}.$$

Assessing balance

The absolute pooled standardized difference between the groups can be calculated for all the covariates using the function MatchBalance in the library Matching.

logistic regression model.

```
library(Matching)
mb <- MatchBalance(qsmk ~ as.factor(sex) + as.factor(race) +
                      age + as.factor(education.code) +
                      smokeintensity + smokeyrs  +
                      as.factor(exercise) +
                      as.factor(active) + wt71, data=nhefshwdat,nboots=1
```

If the absolute value of the standardized mean difference is greater than 10% then this indicates a serious imbalance. For example, sex has an absolute standardized mean difference of $|-16.022| = 16.022$ indicating serious imbalance between the groups in males and females.

Assessing balance in the smoking cessation study

Output from `MatchBalance()`.

```
***** (V3) age *****
before matching:
mean treatment........ 46.174
mean control.......... 42.788
std mean diff......... 27.714

NB: some output is omitted ...
```

Indicates imbalance in distribution of age in smoking and non-smoking group.

If the absolute value of the standardized mean difference is greater than 10% then this indicates a serious imbalance. Age has an absolute standardized mean difference of 46.17 indicating serious imbalance between the groups in age.

Assessing balance in the smoking cessation study

```
***** (V2) as.factor(race)1 *****
before matching:
mean treatment........ 0.08933
mean control.......... 0.14617
std mean diff......... -19.905

mean raw eQQ diff..... 0.057072
med  raw eQQ diff..... 0
max  raw eQQ diff..... 1

mean eCDF diff........ 0.028422
med  eCDF diff........ 0.028422
max  eCDF diff........ 0.056844

var ratio (Tr/Co)..... 0.65287
T-test p-value........ 0.0012863
```

Assessing Balance in the smoking cessation study

```
***** (V14) wt71 *****
before matching:
mean treatment........ 72.355
mean control.......... 70.303
std mean diff......... 13.13

mean raw eQQ diff..... 2.1872
med  raw eQQ diff..... 2.04
max  raw eQQ diff..... 14.75

mean eCDF diff........ 0.032352
med  eCDF diff........ 0.032386
max  eCDF diff........ 0.07

var ratio (Tr/Co)..... 1.0606
T-test p-value........ 0.022421
KS Bootstrap p-value.. 0.1
KS Naive p-value...... 0.10646
KS Statistic.......... 0.07
```

Ignorable Treatment Assignment

What is the effect of Smoking on weight gain?

Outcome = Y = Weight gain

treatment = Smoking = $\begin{cases} 1 & \text{Yes quit} \\ 0 & \text{No quit} \end{cases}$

Treatment assignment T is ignorable if,

$$P(T|Y(0), Y(1), \mathbf{x}) = P(T|\mathbf{x}).$$

Y(0) ~ weight gain when quit = No

Y(1) = weight gain quit = Yes.

Symbolically,

$$T \perp (Y(0), Y(1)) \,|\mathbf{x}.$$

T is conditionally independent of $Y(0), Y(1)$ given covariates $\mathbf{x}$.

Propensity scores and ignorable treatment assignment

Ignorable treatment assignment implies that

$$P(T|Y(0), Y(1), e(\mathbf{x})) = P(T|e(\mathbf{x})),$$

or

$$T \perp Y(0), Y(1)|e(\mathbf{x}).$$

- This means that the scalar propensity score $e(\mathbf{x})$ may be used in place of the many covariates in $\mathbf{x}$.

Propensity scores and ignorable treatment assignment

Ignorable treatment assignment implies that

$$P(T|Y(0), Y(1), e(\mathbf{x})) = P(T|e(\mathbf{x})),$$

or

$$T \perp Y(0), Y(1)|e(\mathbf{x}).$$

- This means that the scalar propensity score $e(\mathbf{x})$ may be used in place of the many covariates in $\mathbf{x}$.
- It may be difficult to find a treated and control unit that are closely matched for every one of the many covariates in x, but it is easy to match on one variable, the propensity score, $e(\mathbf{x})$, and doing that will create treated and control groups that have similar distributions for all the covariates.

Propensity scores and ignorable treatment assignment

- The propensity score can be used in place of many covariates.

Propensity scores and ignorable treatment assignment

- The propensity score can be used in place of many covariates.
- If treatment assignment is strongly ignorable then propensity score methods will produce unbiased results of the treatment effects.

Propensity scores and ignorable treatment assignment

- The propensity score can be used in place of many covariates.
- If treatment assignment is strongly ignorable then propensity score methods will produce unbiased results of the treatment effects.
- In the smoking cessation study what does it mean for treatment assignment to be ignorable?

Propensity scores and ignorable treatment assignment

- The propensity score can be used in place of many covariates.
- If treatment assignment is strongly ignorable then propensity score methods will produce unbiased results of the treatment effects.
- In the smoking cessation study what does it mean for treatment assignment to be ignorable?
- The potential outcomes for weight gain in the smoking cessation (treated) and smoking (control) groups are independent conditional on the propensity score.

Propensity scores and ignorable treatment assignment

- The propensity score can be used in place of many covariates.
- If treatment assignment is strongly ignorable then propensity score methods will produce unbiased results of the treatment effects.
- In the smoking cessation study what does it mean for treatment assignment to be ignorable?
- The potential outcomes for weight gain in the smoking cessation (treated) and smoking (control) groups are independent conditional on the propensity score.
- The treatment assignment mechanism has been reconstructed using the propensity score.

Stop