PS2030
Political Research and Analysis
Unit 4: Models for Causal Inference

1. Counterfactuals, Potential Outcomes, and Causal Inference
2. Selection on Observables: Regression and Matching

Spring 2025, Weeks 12-13

WW Posvar Hall 3600

Professor Steven Finkel

Goals for Session

- Introduction to the "Potential Outcomes" or "Counterfactual" Model of Causality
- Regression Analysis within the Potential Outcomes Framework
- Matching Methods to Handle "Selection on the Observables"

1. The "Potential Outcomes" or Counterfactual Model of Causality

- Developed over the past 60 years and attributed to statisticians Donald Rubin and Paul Holland (and earlier to Jerzy Neyman in the 1920s)
- Virtually all empirical political analysis can in principle be viewed as an attempt to estimate the causal effects of some kind of "treatment" on a particular outcome or set of outcomes
  - Effect of going to college on voting, effect of joining an IO on war, effect of changing electoral laws on number of parties, effects of attending a civic education workshop on political knowledge, etc.
- What do we mean by the "causal effect" of a treatment?
  - Assume a unit can have two "potential" outcomes, depending on whether it is exposed to some treatment *D* or not
    - Y_{1i} is unit *i*'s value of Y if exposed to the treatment (D=1)
    - Y_{0i} is unit *i*'s value of Y if not exposed to the treatment (D=0)
  - So $(Y_{1i} - Y_{0i})$ is the difference in unit *i*'s outcome under the two conditions. It is the difference in the outcome at a given point in time for a unit if it was exposed to D versus the outcome *for that same unit* if it was not exposed to D
  - We call this quantity the *causal effect* of the treatment D

The “Fundamental Problem of Causal Inference”

- Problem: This quantity is *unobservable*!!! We only see one of the two values of Y for a given unit --- Y_{0i} for the control group (D=0), and Y_{1i} for the treatment group (D=1). This is what is known as the **“fundamental problem of causal inference”**!!!
  - We don’t know what the control group would have looked like at a given point in time if it had gotten the treatment ($Y_{1i}|\ D=0$), and we don’t know what the treatment group would have looked like at a given point in time if it had not gotten the treatment ($Y_{0i}|\ D=1$). These “**counterfactual**” outcomes are unobserved, so we cannot directly calculate the causal effect of the treatment.

	Treatment Group	Control Group
Treated	**$Y_{1i}\|D=1$ Observed**	***$Y_{1i}\|D=0$ Counterfactual***
Untreated	***$Y_{0i}\|D=1$ Counterfactual***	**$Y_{0i}\|D=0$ Observed**

- Can formalize this idea as an expression for *observed* Y_i:
- $(1) Y_i = Y_{1i} D_i + Y_{0i}(1 - D_i)$
- which says that we observe *only* the potential outcome associated with the treatment condition for the treatment group, and *only* the potential outcome associated with the control condition for the control group
- Some algebraic manipulation in (1) gives two alternative expressions:
- $2(a)$: $Y_i = Y_{oi} + (Y_{1i} - Y_{0i}) D_i$
- $2(b)$: $Y_i = Y_{1i} - (Y_{1i} - Y_{0i})(1 - D_i)$
- which says a) observed Y is equal to the unit's potential outcome associated with the control condition, plus the unit's treatment effect if it were treated; or b) observed Y is equal to the unit's potential outcome under the treatment condition, minus the unit's treatment effect if it were not treated
- **But we don't observe any unit's treatment effect!!!!**
- Nearly all (modern) empirical social science research is concerned with developing ways of **identifying** and **estimating** the unobservable quantity (Y_{1i} - Y_{0i}).

Treatment Effects and Observed Treatment-Control Group Differences

- We can start with the average difference between treated and control units on the outcome variable: $E(Y_{1i}|D=1) - E(Y_{0i}|D=0)$
- This is ***NOT*** an unbiased estimate of the average treatment effect (ATE)

  $ATE = E(Y_{1i}-Y_{0i}) = E(Y_{1i})-E(Y_{0i})$
- The **"Average Treatment Effect" (ATE)** can be decomposed as:

$$
\begin{aligned}
(3)\quad E(Y_{1i} - Y_{0i}) = D(ATE) &= E(Y_{1i} \mid D=1) - E(Y_{0i} \mid D=0) \\
&- E(Y_{0i} \mid D=1) - E(Y_{0i} \mid D=0) \\
&- (1\text{-}Pr(D=1))*[E(D_i \mid D=1) - E(D_i \mid D=0)]
\end{aligned}
$$

- This says that the **average treatment effect** is composed of three terms:
  - the **observed** difference between average Y for treatment and control units (line 1)
  - the **(unobserved)** difference between what the treatment group *would have looked like* in the absence of treatment and what the control group *did* look like in the absence of treatment (line 2)
  - The **(unobserved)** difference between what the average treatment effect was for treatment units and what the average treatment effect *would have been* for control units, had they been treated (weighted by the proportion of control units) (line 3)

- Alternatively: Observed Differences, Treatment v. Control =

  ATE (Average Treatment Effect) +

  Differences in Potential Control Outcome (Y_{0i}) between Treatment and Control Groups +

  Differences in Treatment Effects between Treatment and Control, weighted by Proportion in Control Group
- This is relevant when you are concerned about "average treatment effects" as a substantively important quantity
- Other potentially important quantities:
  - ATT (Average Treatment Effect on the Treated): what is the effect of treatment on units that take up the treatment?
  - ATC (Average Treatment Effect on Control): what is the effect of treatment on units that do not take up the treatment, but theoretically *could* take up the treatment?
- These quantities could be different ("heterogeneous" treatment effects); we'll often simplify and assume ATE=ATT=ATC. Then line 3 in equation (3) would drop out.

- We observe the difference between the treatment group and the control group: $(Y_{1i}|D=1) - (Y_{0i}|D=0)$
- If a (potential) causal effect is constant for treatment and control groups, then, following equation (3):

$$E((Y_{1i}|D=1) - (Y_{0i}|D=0)) = E((Y_{1i}|D=1) - (Y_{0i}|D=1)) + E((Y_{0i}|D=1) - (Y_{0i}|D=0))$$

Observed Group Difference = ***"Treatment Effect"*** + ***"Baseline Selection Bias"***

$$E((Y_{1i}|D=1) - (Y_{0i}|D=0)) = E((Y_{1i}|D=1) - (Y_{0i}|D=1)) \quad + \quad E((Y_{0i}|D=1) - (Y_{0i}|D=0))$$

Observed Difference = “Treatment Effect” + “Baseline Selection Bias”

- This means that the observed difference between the treatment and control groups is a function of the (unobserved) causal effect of the treatment on units that get the treatment PLUS the (unobserved) difference in the “no-treatment” outcome between the treatment and control groups.
- The latter term is the difference in what the treatment group ***would have looked like*** in the absence of treatment and what the control group ***did look like*** in the absence of treatment
- Whenever the “selection bias” term is zero, or whenever $E(Y_{0i}|D=1) = E(Y_{0i}|D=0)$, then observed differences between treatment and control groups = the causal effect of the treatment

Randomization Solves the Selection Bias Problem!

- When the “baseline selection bias” term is zero, i.e. when $E(Y_{0i}|D=1) = E(Y_{0i}|D=0)$, then observed differences between treatment and control groups will equal the causal effect of the treatment
- When will this occur? If treatment and control groups are **randomly assigned**, then their baseline potential “non-treatment” outcomes are equalized. The observed control group outcome will be (statistically) the same as what the observed treatment group’s outcome *would have looked like* had it not received the treatment. (And the observed treatment group’s outcome will be (statistically) the same as what the observed control group’s outcome *would have looked like* had it actually received treatment too)
- In other words, under randomization:

  $\mathbf{E(Y_{0i}|D=1)} = E(Y_{0i}|D=0)$ and $E(Y_{1i}|D=1) = \mathbf{E(Y_{1i}|D=0)}$

  with counterfactual potential outcomes in boldfaced type
- We can say that, under randomization, treatment status and potential outcomes are independent; treatment assignment is “ignorable”:

  $Y_0, Y_1 \perp D$

- How does this work? Randomization equates the treatment and control groups on ***all*** variables --- *both observed and unobserved* --- that could have produced observed baseline differences between the groups.
- This means that we can identify the ATE easily, since randomization guarantees first, that there is no baseline "selection bias" differences between treatment and control groups, i.e., the control group's outcome is exactly the same as what the treatment group's outcome would have been in the absence of treatment; and
- Second, the expected treatment effect for treated units will be the same as the expected treatment effect for control units would have been had they been treated. So equation 3 (slide 6) reduces to a comparison of treatment and control group means, which is the ATE as well as the ATT and the ATC
- This is the beauty of random assignment for causal inference! We can use the control group mean as a **perfect proxy** for what the treatment group mean would have been in the absence of treatment; and we can use the treatment group mean as a **perfect proxy** for what the control group mean would have been had it received treatment. Thus randomization is a very attractive identification strategy if it is possible to implement!

Selection Bias in Observational Research

- In non-experimental or “observational” research, we face the ever-present possibility (probability) of selection bias, that the treatment and control groups will have different baseline (non-treatment) outcomes due to pre-existing differences on relevant *observed* and/or *unobserved* variables
- Different methods exist for attempting to identify causal effects in the presence of these biases, some of which we have covered already but not using the same terminology or formal framework for causal inference
- Methods for controlling for selection biases due to observed variables:
  - Multivariate regression and “regression adjustment”
  - “Matching” and “Propensity Score Matching”
- Methods for controlling for selection biases due to *unobserved* variables:
  - Instrumental Variables and “Natural Experiments”
  - “Difference in Difference” and longitudinal panel data models
  - Heckman selection models
  - Regression Discontinuity Designs (RDD)
- “We will only have time to discuss some of these methods”
- For more, see PS2702 Causal Inference and PS2701 Longitudinal Analysis

Example: Potential Outcomes in Non-Experimental Research

Unit	X	Y0	Y1	D	Y	Treatment Effect	Treatment Effect, D=1	Treatment Effect, D=0
1	0	1	0	1	0	-1	-1	
2	0	0	0	0	0	0		0
3	0	0	0	0	0	0		0
4	0	0	0	0	0	0		0
5	1	0	14	0	0	14		14
6	1	0	14	0	0	14		14
7	1	0	14	1	14	14	14	
8	1	0	14	1	14	14	14	
MEAN		0.125	7	0.375	3.5	6.875	9	5.6

- $E(Y|D=1)= 9.33$ $E(Y|D=0)=0$ so Observed Difference between Treatment/Control=9.33
- Colored cells are **unobserved counterfactuals**
- **TRUE ATE**$=E(Y1-Y0)=E(Y1)-E(Y0)=6.875$
- **TRUE ATT**$=E(Y1-Y0|D=1)=9.00$
- Baseline Selection Bias: $E(Y0|D=1)-E(Y0|D=0)=.3333$
- Differential Treatment Effect for D=1 and D=0, is 3.4 (since ATT=9.0 and ATC=5.6)
- **ATE**=(Observed Difference-Baseline Selection Bias-(1-P(D=1)*Differential Treatment Effect, D=1, D=0)=9.333-.333-(.625*3.4)=6.875
- **ATT**=(Observed Difference-Baseline Selection Bias)=9
- **Problem: All of these effects are unobservable! We only observe Y, D, and X! How can we identify and estimate the ATE and/or ATT given the observed data?**

Selection on the Observables

- Selection bias will not be a problem whenever *all* relevant variables that are responsible for the baseline potential "non-treatment" outcome differences are measured and incorporated into the model: $E(Y_{0i}|D=1, X_i) = E(Y_{0i}|D=0, X_i)$
- If we can assume that the Xs that would equate baseline "non-treatment outcomes" are **observed** variables, we call this "*selection on the observables*", or that treatment assignment is "ignorable given the Xs": $Y_0, Y_1 \perp D|X$.
- This is the "identifying assumption" that allows estimation of the causal effects of interest (and which certainly could be wrong).
  - It means that we attempt to make the treatment and control groups equal on potential outcomes Y_0 and Y_1 by balancing the groups on "confounding" X variables, and then we observe differences in Y among the treatment and control groups that have been equated on X.
  - In regression, we "adjust" for confounders, variables that may also determine Y
  - In matching, we balance on confounders that may also determine D, and we say that after balancing on the Xs, treatment assignment is "as good as random".

- In multiple regression, we attempt to ensure that $E(Y_{0i}|D=1)=E(Y_{0i}|D=0)$ by "controlling" or "adjusting" for all observable X that may be related to both D and Y and bringing them into the equation.
- Assuming constant effects of X on Y in both treatment and control groups, we obtain the familiar regression set up:
- $Y_i = \alpha_0 + \rho D_i + \beta X_i$

- We can also express this equation in terms of average differences between treatment and control groups:
- $E(Y_i|D = 1) = \alpha_0 + \rho + \beta \bar{X}_{D=1}$
- $E\ (Y_i|D = 0) = \alpha_0 + \quad\quad \beta \bar{X}_{D=0}$
- $E(Y_i|D = 1) - (Y_i|D = 0) = \rho + \beta(\bar{X}_{D=1} - \bar{X}_{D=0})$

This is *exactly* what we did when we examined dummy variable regression!! (In fact, this procedure is called "covariance or **regression adjustment**"). If baseline selection bias is controlled by bringing X into the analysis in this fashion (i.e., if the identifying assumptions of non-ignorable treatment assignment and other assumptions to be noted are correct), then ρ is the causal effect of the treatment! That is multiple regression!

Regression Adjustment in Practice

- This also suggests a more elaborate method of estimating causal effects via regression adjustment
- Allow separate regressions for treatment and control observations, as we did with Causal Mediation Analysis
- Then, for each unit, calculate **a predicted outcome, given treatment** from their values on the covariates using the treatment group regression model, and a **a predicted outcome, given control**, using the control group regression model
- $(Y_i|D = 1) = \alpha_0 t + \beta_t X$
- $(Y_i|D = 0) = \alpha_0 c + \beta_c X$

- Use these two predicted outcomes as estimates for each individuals' *potential outcomes* $\mathbf{Y_{1i}}$ **and** $\mathbf{Y_{0i}}$
- Take the difference for each individual, and average across the sample

Regression Adjustment Example

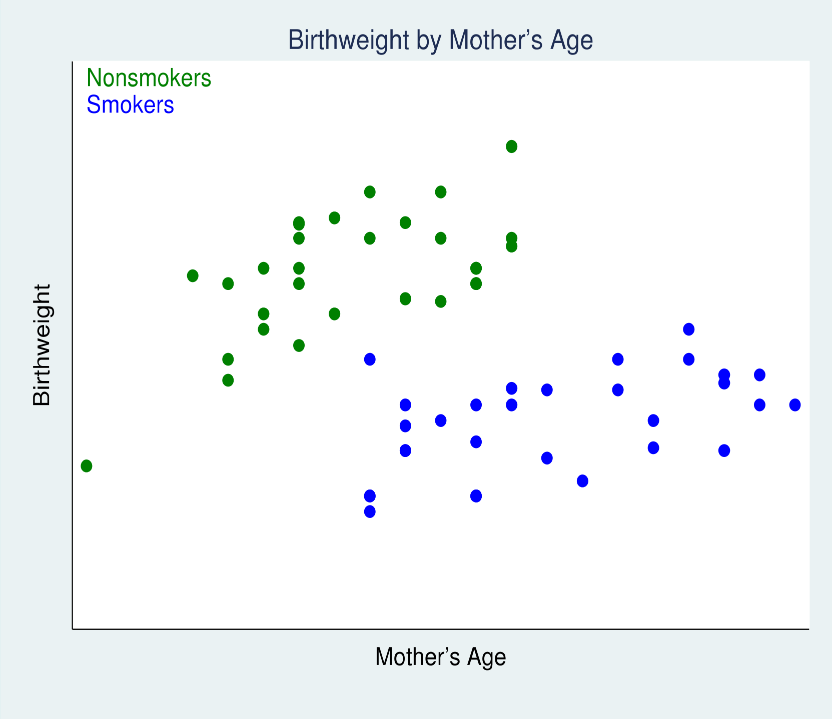
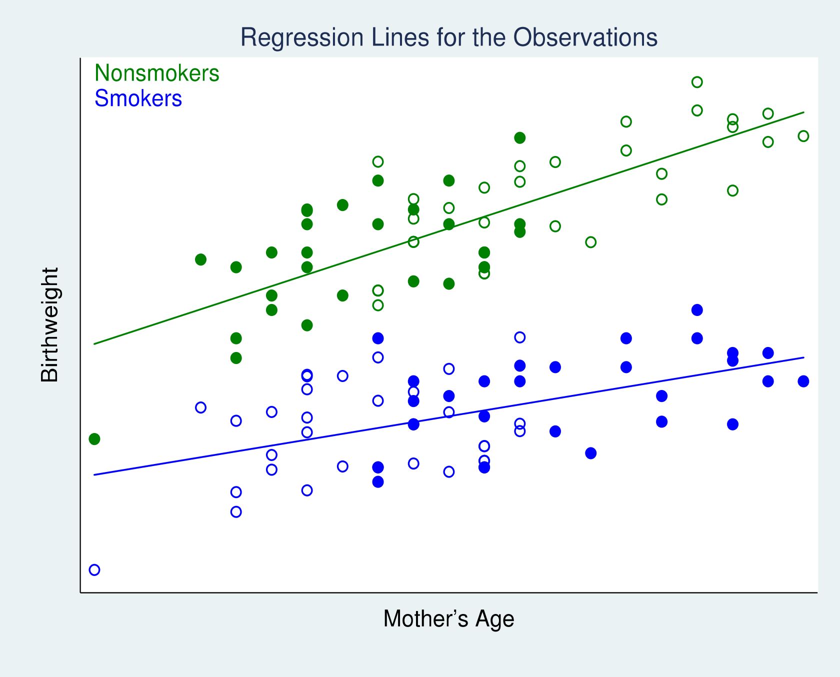


Figure 3



Our Example

Unit	X	Y0	Y1	D	Y	Treatment	Treatment	Treatment
						Effect	Effect, D=1	Effect, D=0
1	0	1	0	1	0	-1	-1	
2	0	0	0	0	0	0		0
3	0	0	0	0	0	0		0
4	0	0	0	0	0	0		0
5	1	0	14	0	0	14		14
6	1	0	14	0	0	14		14
7	1	0	14	1	14	14	14	
8	1	0	14	1	14	14	14	
MEAN		0.125	7	0.375	3.5	6.875	9	5.6

TRUE ATE=E(Y1-Y0)=E(Y1)-E(Y0)=6.875

Mean X, Treatment=.67
Mean X, Control= .40

Does regression or regression adjustment recover the causal effect?

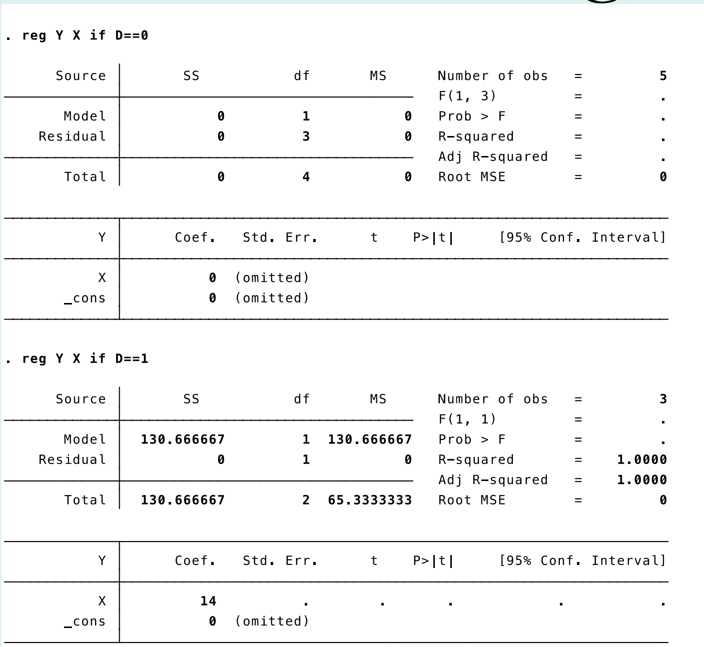
REGRESSION OF Y ON D: $\beta_1 = 9.33$ WRONG!

REGRESSION OF Y ON D, X: $\beta_1 = 8.00$ $\beta_2 = 5$ WRONG!

$(9.33-5*.2667)=8$

```
reg Y D

      Source |       SS       df       MS              Number of obs =       8
-------------+------------------------------           F(  1,     6) =    7.50
       Model |  163.333333     1  163.333333           Prob > F      =  0.0338
    Residual |  130.666667     6  21.7777778           R-squared     =  0.5556
-------------+------------------------------           Adj R-squared =  0.4815
       Total |         294     7          42           Root MSE      =  4.6667

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   9.333333   3.408051     2.74   0.034     .9941318    17.67253
       _cons |          0   2.086997     0.00   1.000    -5.106697    5.106697
------------------------------------------------------------------------------

. reg Y D X

      Source |       SS       df       MS              Number of obs =       8
-------------+------------------------------           F(  2,     5) =    6.25
       Model |         210     2         105           Prob > F      =  0.0436
    Residual |          84     5        16.8           R-squared     =  0.7143
-------------+------------------------------           Adj R-squared =  0.6000
       Total |         294     7          42           Root MSE      =  4.0988

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |          8   3.098387     2.58   0.049     .0353435    15.96466
           X |          5          3     1.67   0.156    -2.711746    12.71175
       _cons |         -2    2.19089    -0.91   0.403    -7.631863    3.631863
------------------------------------------------------------------------------
```

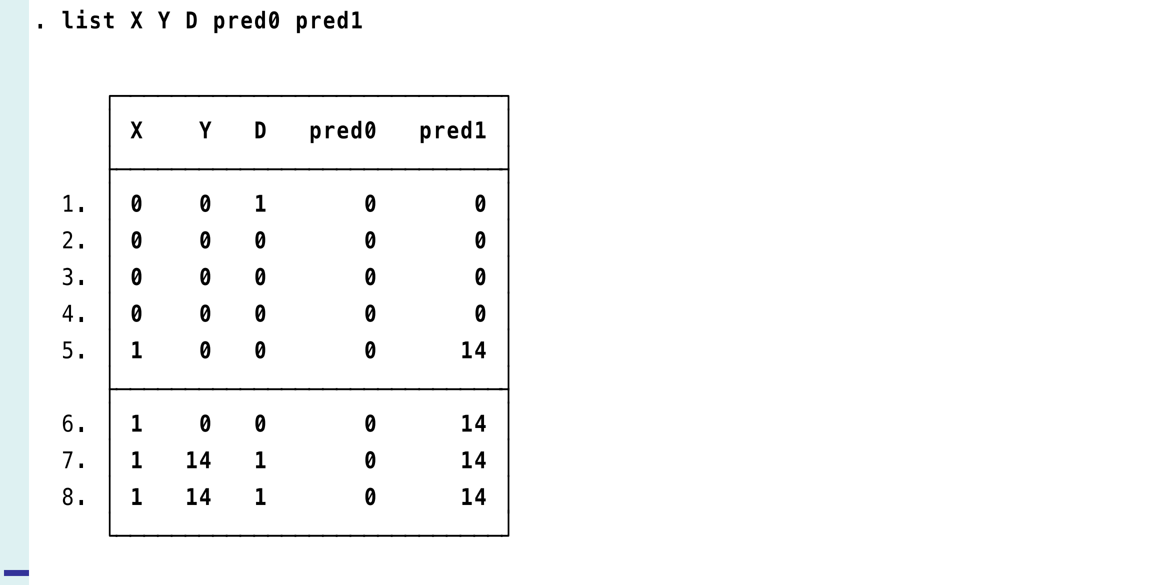
Regression Adjustment



```
. reg Y X if D==0

      Source |       SS           df       MS      Number of obs   =         5
-------------+----------------------------------   F(1, 3)         =         .
       Model |          0         1           0   Prob > F        =         .
    Residual |          0         3           0   R-squared       =         .
-------------+----------------------------------   Adj R-squared   =         .
       Total |          0         4           0   Root MSE        =         0

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |          0  (omitted)
       _cons |          0  (omitted)
------------------------------------------------------------------------------

. reg Y X if D==1

      Source |       SS           df       MS      Number of obs   =         3
-------------+----------------------------------   F(1, 1)         =         .
       Model | 130.666667         1  130.666667   Prob > F        =         .
    Residual |          0         1           0   R-squared       =    1.0000
-------------+----------------------------------   Adj R-squared   =    1.0000
       Total | 130.666667         2  65.3333333   Root MSE        =         0

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |         14          .        .       .            .           .
       _cons |          0  (omitted)
------------------------------------------------------------------------------
```

Regression slope of Y on X in control group=0
Regression slope of Y on X in treatment group =14

Predicted Y in control condition: 0
Predicted Y in treatment condition:
0 if X is 0; 14 if X is 1



```
. list X Y D pred0 pred1

     +-------------------------------+
     | X    Y   D   pred0    pred1   |
     |-------------------------------|
  1. | 0    0   1       0        0   |
  2. | 0    0   0       0        0   |
  3. | 0    0   0       0        0   |
  4. | 0    0   0       0        0   |
  5. | 1    0   0       0       14   |
     |-------------------------------|
  6. | 1    0   0       0       14   |
  7. | 1   14   1       0       14   |
  8. | 1   14   1       0       14   |
     +-------------------------------+
```

So: Average Treatment Effect (ATE)=
(4*0 + 4*14) /8 = 7.0

Average Treatment Effect on Treated (ATT)= 28/3=9.333

Average Treatment Effect on Control (ATC)= 28/5=5.60

Problems with Regression and Regression Adjustment

- The *unobservables* problem: regression and regression adjustment are not able to handle selection biases due to *unobservables*
- The *functional form* problem: Regression assumes that the Xs enter the equation in a linear, additive fashion and that the data are drawn from a probability distribution of a given form (usually "normal"). This may not be the case. X could affect both D and Y in any number of ways that do not follow an easily predetermined form, and imposing the form on slide 16 may not fully "control" for X in the estimation of the treatment effect ρ.
  - For example, if exposure to civic education is D, knowledge is Y and group memberships is X, can we assume that Y is necessarily a linear function of X? Maybe threshold of 2 groups is necessary, or step effects such that 5-6 groups really adds much more knowledge than 1 or 2? If so, then multiple regression and regression adjustment will not sufficiently control for X!

- The *common support* problem: Regression (multiple regression (and regression adjustment) says that, at every level of $X_{i,}$ we will see a difference on Y between treatment and control units. But there is no guarantee, using multiple regression, that there will be *any* comparable control units for all treated units at each level of X_i
  - In fact, the larger the difference between the means of the covariate X for the treatment and control groups, the more that we may be extrapolating beyond the region of common support by using regression (see the birthweight example)
- The *functional form problem* and the *common support problem* are the major motivations behind the use of "matching" methods for causal inference as opposed regression!

Matching

- Basic idea: match each treatment unit *i* with a control unit *j* that has the same characteristics on X, and calculate the average differences in Y for all of these "matched pairs" to estimate the causal effect of interest (ATT, ATC, ATE)
- For example:
- $\Delta\left(ATT_{Matching}\right) = \frac{1}{N_{D=1}} \sum\left[\left(Y_{1i} \middle| D = 1, X = x\right) - \left(Y_{0j} \middle| D = 0, X = x\right)\right]$
- Matching allows the relationship between X, D and Y to have any kind of functional forms. Whatever the form is, we find matches between treatment and control units at a given level of X and use that control unit as the counterfactual "no-treatment" proxy for the treated unit
- Matching (in this fashion) also ensures common support; if no suitable control unit is found for a given treated unit at some level of X, the treated unit can be discarded and not considered further
- Assumption for $\Delta ATT_{Matching}$ to correctly estimate the "true" ATT? As in regression, that treatment assignment is "ignorable given the Xs": $Y_0 \perp D | X$ (for ATT; $Y_0, Y_i \perp D | X$ for ATE)

- Most straightforward approach to matching: find the control unit that *exactly* matches the treated unit on *all* relevant covariates. So, match each treated male with a control male, each treated female with a control female, each treated young male with a control young male, each treated young minority male with a control young minority male, etc.
- Big problem: What happens as multiple X variables determine D? How do we find exact matches for a treatment unit with values $X_i=x_i$ for **all** of the Xs? If D is exposure to civic education, for example, can we be certain that we will have a highly interested, high media exposure, urban, young minority male who did not attend a workshop for a given treated individual with those same characteristics? If not, we have no counterfactual control unit for that treatment person. As more and more Xs affect D, dimensionality problem becomes even more acute.
- If there are 20 X variables, for example, and even if each of them is dichotomous, there will be 2^{20} , or 1,048,576 possible cells or combinations of the Xs,
- Breakthrough: Rosenbaum/Rubin's 1980s work on **"Propensity Score Matching"**
- **PSM**: constructs a comparison group by modeling the *probability* of all units being treated on the basis of a full set of observed characteristics, and then matches each treatment unit with the control unit or units that (counterfactually) had the closest probability of being treated for purposes of estimating the ATT
- Rosenbaum and Rubin show that matching on P(D=1) is as good as matching on **X**

Propensity Score Matching

- The propensity score model, given *exact* matching of treated case i with control case j on P(D=1 | X_i):
- $\Delta(ATT_{PSM}) = \frac{1}{N_{D=1}} \sum \left[(Y_{1i} | D = 1, P(Y = 1|X) = x) - \left(Y_{0j} \middle| D = 0, P(Y = 1|X) = x \right) \right]$
- There are many alternative ways of aggregating and weighting control unit(s) to arrive at the best counterfactual for a given treatment unit, so a more general way to express the PSM model is:
- $\Delta(ATT_{PSM})=\frac{1}{N_{D=1}} \sum \left[(Y_{1i} | D = 1, P(Y = 1|X) = x) - \left(\omega Y_{0j} \middle| D = 0, P(Y = 1|X) = x \right) \right]$
- where ω is the weighting mechanism for aggregating the control unit observation(s) that will serve as the match for a given treatment unit
- Identifying assumption, as in regression and exact matching:
  - Conditional on the Xs, treatment assignment is ignorable, or "as good as random".
  - Assumption of $Y_0 \perp D | X$ for ATT, $Y_0, Y_1 \perp D | X$ for ATE

Steps in Propensity Score Matching

1. Estimate propensity score $P(D=1|X)$ for all units using logit/probit
2. Define the region of "common support", i.e., the region where the distributions of propensity scores for treatment and control units overlap
3. Construct a matched treatment and control sample using the propensity score in one of several ways (nearest neighbor, etc.). This corresponds to the ω on the previous slide
4. Conduct tests of "balance" between the treatment and control groups on all X variables. If treatment assignment is ignorable, given the propensity score, then treatment and control groups should have nearly identical levels of Xs at similar values of the propensity score. If not, repeat steps 1-3.
5. Conduct post-matching analysis to arrive at ATT (or ATE). This can be done in a variety of ways:
   - Calculate average differences on Y between the matched treatment and control pairs
   - Stratify the sample into blocks on the propensity score, estimate effects within blocks using regression etc., aggregate via weighting by the size of the blocks
   - Weighted regression, with the inverse of the propensity score serving as the weights

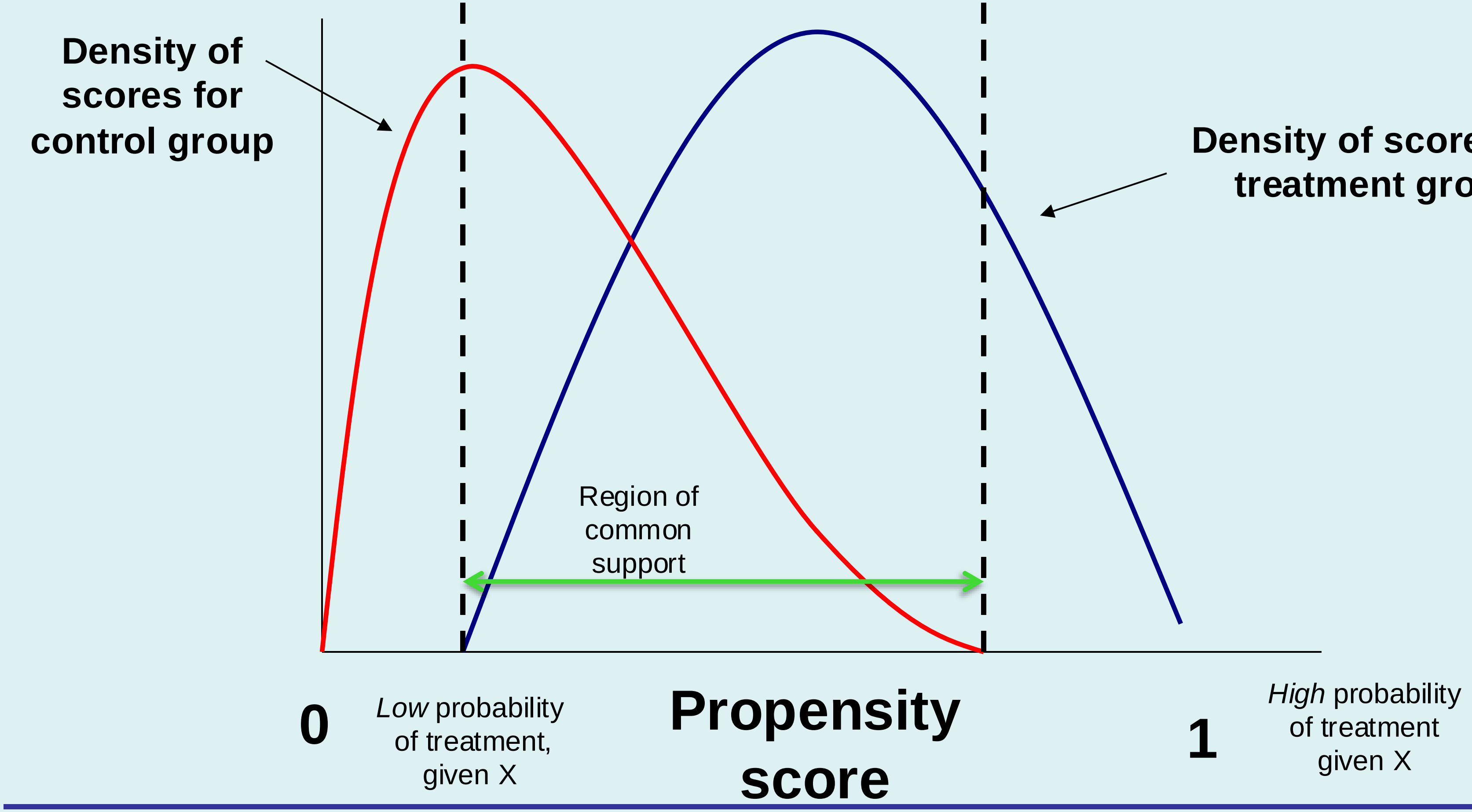
Step 1: Estimate the Propensity Score

- Run a regression (probit or logit) that has the probability of receiving treatment on the left hand side, and the covariates that determine selection into the treatment on the right hand side (i.e., XB):
- $P(D = 1) = \Phi(XB)$ probit

  $P(D = 1) = \frac{exp^{XB}}{1+exp^{XB}}$ logit
- The propensity score is just the predicted probability of D=1 that you get from this regression. (Some methods recommend using the linear term (either the z-score or the logit)).
- Start with a simple specification (e.g. just linear terms). Then, depending on the remaining imbalance on the Xs that exists once you compare treatment and controls at similar levels of P(D=1), modify the propensity score regression by including squared and/or interaction terms, or new variables.
- You want parsimony but also a model that fully satisfies the requirements of "ignorable treatment assignment" once the propensity score is controlled. Be relatively liberal in including (pre-treatment) covariates.

Step 2: Define the Region of Common Support

Once the propensity score is calculated, eliminate control units with lower propensities to be treated than the lowest treated unit, and eliminate treatment units with higher propensities to have been treated than the highest control unit


Density of scores for control group
Density of scores for treatment group
Region of common support
0
Low probability of treatment, given X
Propensity score
1
High probability of treatment given X

Step 3: Construct Matched Treatment-Control Sample

1. Nearest Neighbor Matching
   - Select for each treated individual *i* the control individual *j* with the smallest propensity score distance from individual. Then discard cases *i (*and possibly *j*) from further consideration, repeat until all treated units have a matched *j* control

- $\Delta(ATT_{PSM}) = \frac{1}{N_{D=1}} \sum\left[(Y_{1i}|D = 1, P(Y = 1)X = d) - \left(\omega Y_{0j}\middle|D = 0, P(Y - 1)X = d\right)\right]$

   - where $\omega = 1$ if $\left|P_i - P_j\right| = min_{kD=0}|P_i - P_k|$; 0 otherwise

   - If $P_i = P_j$, nearest neighbor matching is *exact* matching on the propensity score
   - Can potentially discard many cases if control units not needed for given matches (but maybe not a real problem given the increased precision of the ATT estimates)
   - Can lead to some very poor matches (since “nearest” doesn’t mean “near”)
   - Can also specify *“k:1”* matching where multiple control units (N=k) can serve as matches for a given treatment unit.
   - Can match without replacement, or “with replacement” so control units can serve as matches for multiple treatment cases; this complicates standard errors

2. Propensity Score Weighting

- An alternative and very commonly utilized procedure is to use the propensity scores as weights, similar to sampling weights in regular regression or other analyses.
- For the ATE, define the weights as:
- $\omega_{treatment} = \frac{1}{P(D=1)}$ and $\omega_{control} = \frac{1}{1-P(D=1)}$
- Treated units with high P(D=1) are down-weighted, control units with high P(D=1) are up-weighted
- For the ATT, define the weights as:
- $\omega_{treatment} = 1$ and $\omega_{control} = \frac{P(D=1)}{1-P(D=1)}$
- Control units with high P(D=1) upweighted via the denominator, then matched to treatment group via the numerator
- Advantages: Easy to implement and uses all the cases in the sample!

Step 4: Check for Covariate Balance in the Matched Sample

- Whatever method is chosen to produce the matched sample, need to check on the quality of the matches by examining **covariate balance between the treated and control units** before and after the matches were constructed.
- If successful, matching on the propensity score should produce a matched sample where the treated and control units have similar distributions on *all* covariates. In that case, we are closer to fulfilling the "ignorable treatment assignment" assumption, in that, given the propensity score, treatment assignment is "as good as random" (at least on the observable covariates included in the propensity score calculation)
- If unsuccessful, need to re-specify the propensity score and/or the matching method and/or add the covariate to the statistical analysis later on to further "control" for its effects, over and above its inclusion in calculating the propensity score

- Simple way to do this is to conduct t-tests of means pre/post matching and see whether the matched sample is completely balanced. Commonly utilized, though some argue that balance is not really a statistical inference issue – it is a sample property, so t-tests of significance are not technically relevant. Also, since we are potentially throwing out a lot of data in the matching process, we lose statistical power for such tests. This controversy is not yet resolved.
- Rosenbaum and Rubin (1985) recommend a related measure, the *standardized difference in means* or "*standardized bias*" for the treatment and control groups:
- $\frac{\bar{x}_t - \bar{x}_c}{\sigma_t}$
- There should be *no* standardized biases greater than .25, and ideally there should be at least 95% reduction from the pre-matching levels
- Note: need to use *same* matching scheme in balance checking as was done to create the matches in the first place

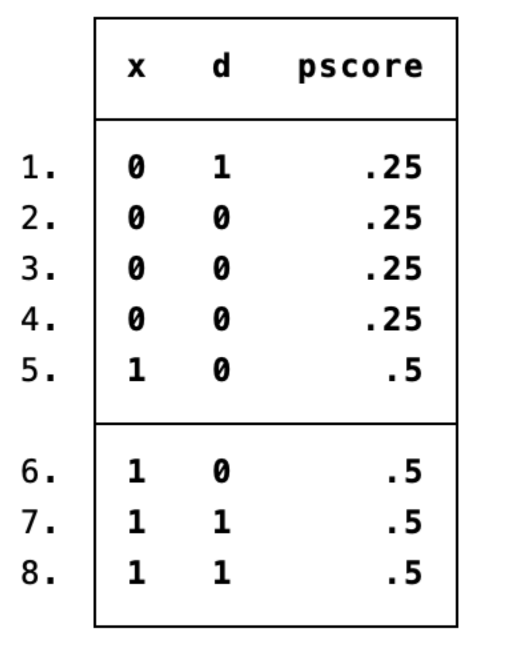
Step 5: Post-Matching Analysis

- Once a successful matched (or weighted matched) sample has been created, the analysis of the outcome can then proceed.
- After nearest neighbor matching: pooled t-test between treatment and control groups in the matched sample on the outcome variable; t-test between matched treatment/control group pairs on the outcome; regression of outcome variable on treatment, controlling for the propensity score and/or any X covariates which failed the balance test in step 4.
- After propensity score weighting, estimate regression models with the weights calculated above used in same way as a sampling weight
  - ATE Estimate for our hypothetical example using weighting?
    - β_1= 7.00 (True=6.875)
  - Standard errors in all weighting models subject of much debate – how much do we take the uncertainty in the propensity score estimation and matches into account? We won't go into these controversies
  - STATA modules for propensity score analysis:

    teffects psmatch, teffects ipw for inverse probability weighting

Example with Propensity Score Weighted Regression

```
logit D X
predict pscore
g ipw=1/(pscore) if D==1
replace ipw=1/(1-pscore) if D==0
reg Y D [weight=ipw]
```



```
     +-----------------+
     | x   d   pscore  |
     |-----------------|
  1. | 0   1      .25  |
  2. | 0   0      .25  |
  3. | 0   0      .25  |
  4. | 0   0      .25  |
  5. | 1   0       .5  |
     |-----------------|
  6. | 1   0       .5  |
  7. | 1   1       .5  |
  8. | 1   1       .5  |
     +-----------------+
```

```
reg Y D [weight=ipw]
(analytic weights assumed)
(sum of wgt is   1.6000e+01)

      Source |       SS       df       MS              Number of obs =       8
-------------+------------------------------           F(  1,     6) =    3.00
       Model |          98     1          98           Prob > F      =  0.1340
    Residual |  195.999999     6  32.6666664           R-squared     =  0.3333
-------------+------------------------------           Adj R-squared =  0.2222
       Total |  293.999999     7  41.9999998           Root MSE      =  5.7155

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |          7   4.041452     1.73   0.134    -2.889076    16.88908
       _cons |  -8.88e-16   2.857738    -0.00   1.000    -6.992633    6.992633
------------------------------------------------------------------------------
```

TRUE ATE=E(Y1-Y0)=E(Y1)-E(Y0)=6.875

- All of these models have been developed for *dichotomous* treatments. Recent work extends these ideas to cases where the treatment variable is ordinal or continuous . See Yanovitsky *et al.* (2005) and Imai and van Dyk (2004) for alternatives based on multiple propensity scores for each category, ordinal logit/probit propensity score weighting, and other models.
- All of these models – indeed, all of the discussion of causal inference so far, depends on one additional assumption known as the Stable Unit Treatment Value assumption (SUTVA). SUTVA means that the potential outcomes of one unit are unaffected by the treatment status of other units. So my potential outcome under treatment (Y_1) or control (Y_0) does not depend on whether *you* or anyone else has been assigned to treatment or control. This would be violated in cases where, for example, there are "spillover effects" of treatment, such that having a lot of treated units in a given area or social network affects the potential outcomes of those untreated units. Much work is currently being done on the estimation of causal effects with "interference" between units; see one influential political science treatment in Sinclair, McConnell, Green (*AJPS* 2012).