Addressing Alternative Explanations: Multiple Regression

17.871
Did Clinton hurt Gore example

- Did Clinton hurt Gore in the 2000 election?
  - Treatment is not liking Bill Clinton
- How would you test this?
Bivariate regression of Gore thermometer on Clinton thermometer
Did Clinton hurt Gore example

- What alternative explanations would you need to address?
- Nonrandom selection into the treatment group (disliking Clinton) from many sources
- Let’s address one source: party identification
- How could we do this?
  - Matching: compare Democrats who like or don’t like Clinton; do the same for Republicans and independents
  - Multivariate regression: control for partisanship statistically
Democratic picture
Independent picture

Gore thermometer vs. Clinton thermometer
Republican picture

Gore thermometer

Clinton thermometer
Combined data picture

Gore thermometer

Clinton thermometer
Combined data picture with regression: bias!
Combined data picture with “true” regression lines overlaid
Tempting yet wrong normalizations

Subtract the Gore therm. from the avg. Gore therm. score

Subtract the Clinton therm. from the avg. Clinton therm. score
3D Relationship
The Linear Relationship between Three Variables

\[ Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i \]
Multivariate slope coefficients

Bivariate estimate: \[ \hat{\beta}_1^B = \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)} \] vs.

Multivariate estimate: \[ \hat{\beta}_1^M = \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)} - \hat{\beta}_2^M \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1)} \]

When does \( \hat{\beta}_1^B = \hat{\beta}_1^M \)? Obviously, when \( \hat{\beta}_2^M \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1)} = 0 \)

Clinton effect (on Gore) in bivariate (B) regression

Party ID effect (on Gore) in multivariate (M) regression

Clinton effect on Party ID in bivariate regression

Clinton effect on Gore in multivariate regression
The Slope Coefficients

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (\bar{Y} - Y_i)(\bar{X}_1 - X_{1,i})}{\sum_{i=1}^{n} (\bar{X}_1 - X_{1,i})^2} - \hat{\beta}_2 \frac{\sum_{i=1}^{n} (\bar{X}_1 - X_{1,i})(\bar{X}_2 - X_{2,i})}{\sum_{i=1}^{n} (\bar{X}_1 - X_{1,i})^2}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} (\bar{Y} - Y_i)(\bar{X}_2 - X_{1,i})}{\sum_{i=1}^{n} (\bar{X}_2 - X_{2,i})^2} - \hat{\beta}_1 \frac{\sum_{i=1}^{n} (\bar{X}_1 - X_{1,i})(\bar{X}_2 - X_{2,i})}{\sum_{i=1}^{n} (\bar{X}_2 - X_{2,i})^2}$$

and

$X_1$ is Clinton thermometer, $X_2$ is PID, and $Y$ is Gore thermometer
The Slope Coefficients More Simply

\[
\hat{\beta}_1 = \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)} - \hat{\beta}_2 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1)} \quad \text{and} \\
\hat{\beta}_2 = \frac{\text{cov}(X_2, Y)}{\text{var}(X_2)} - \hat{\beta}_1 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_2)}
\]

X_1 is Clinton thermometer, X_2 is PID, and Y is Gore thermometer
The Matrix form

\[
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}
= \begin{bmatrix}
1 & x_{1,1} & x_{2,1} & \cdots & x_{k,1} \\
1 & x_{1,2} & x_{2,2} & \cdots & x_{k,2} \\
1 & \cdots & \cdots & \cdots & \cdots \\
1 & x_{1,n} & x_{2,n} & \cdots & x_{k,n}
\end{bmatrix}
\]

\[
\beta = (X'X)^{-1} X'y
\]
3D Linear Relationship
The Output

```
.reg gore clinton party3
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 1745</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>629261.91</td>
<td>2</td>
<td>314630.955</td>
<td>F( 2, 1742) = 1048.04</td>
</tr>
<tr>
<td>Residual</td>
<td>522964.934</td>
<td>1742</td>
<td>300.209492</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>1152226.84</td>
<td>1744</td>
<td>660.68053</td>
<td>R-squared = 0.5461</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.5456</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 17.327</td>
</tr>
</tbody>
</table>

|              | Coef.         | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------------|---------------|-----------|------|------|----------------------|
| gore         |               |           |      |      |                      |
| clinton      | .5122875      | .0175952  | 29.12| 0.000| .4777776 .5467975   |
| party3       | 5.770523      | .5594846  | 10.31| 0.000| 4.673191 6.867856   |
| _cons        | 28.6299       | 1.025472  | 27.92| 0.000| 26.61862 30.64119   |

**Interpretation of clinton effect:** Holding constant party identification, a one-point increase in the Clinton feeling thermometer is associated with a .51 increase in the Gore thermometer.
Separate regressions

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>23.1</td>
<td>55.9</td>
<td>28.6</td>
</tr>
<tr>
<td>Clinton</td>
<td>0.62</td>
<td>--</td>
<td>0.51</td>
</tr>
<tr>
<td>Party</td>
<td>--</td>
<td>15.7</td>
<td>5.8</td>
</tr>
</tbody>
</table>
Is the Clinton effect causal?

- That is, should we be convinced that negative feelings about Clinton really hurt Gore?

- No!
  - The regression analysis has only ruled out nonrandom selection on party ID.
  - Nonrandom selection into the treatment could occur from
    - Variables other than party ID, or
    - Reverse causation, which is feelings about Gore influencing feelings about Clinton.
  - Additionally, the regression analysis may not have entirely ruled out nonrandom selection on party ID because it may have assumed he wrong functional form.
    - E.g., what if nonrandom selection on strong Republican/strong Democrat
Summary: Why we control

- Address alternative explanations by removing confounding effects
- Improve efficiency
Why did the Clinton Coefficient change from 0.62 to 0.51

```
corr gore clinton party, cov
(obs=1745)
```

<table>
<thead>
<tr>
<th></th>
<th>gore</th>
<th>clinton</th>
<th>party3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gore</td>
<td>660.681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clinton</td>
<td>549.993</td>
<td>883.182</td>
<td></td>
</tr>
<tr>
<td>party3</td>
<td>13.7008</td>
<td>16.905</td>
<td>.8735</td>
</tr>
</tbody>
</table>
The Calculations

\[ \hat{\beta}_1^B = \frac{\text{cov}(\text{gore, clinton})}{\text{var}(\text{clinton})} = \frac{549.993}{883.182} = 0.6227 \]

\[ \hat{\beta}_1^M = \frac{\text{cov}(\text{gore, clinton})}{\text{var}(\text{clinton})} - \hat{\beta}_2^M \frac{\text{cov}(\text{clinton, party})}{\text{var}(\text{clinton})} \]

\[ = \frac{549.993}{883.182} - 5.7705 \frac{16.905}{883.182} \]

\[ = 0.6227 - 0.1105 \]

\[ = 0.5122 \]
Accounting for total effects

$$\hat{\beta}_1^M = \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)} - \hat{\beta}_2 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1)}$$

(i.e., regression coefficient when we regress $X_2$ (as dep. var.) on $X_1$ (as ind. var.))

$$\hat{\beta}_1^M = \hat{\beta}_1^B - \hat{\beta}_2^M \gamma_{21}^M$$

$$\hat{\beta}_1^B = \hat{\beta}_1^M + \hat{\beta}_2^M \gamma_{21}^M$$
Accounting for the total effect

\[ \hat{\beta}_1^B = \hat{\beta}_1^M + \hat{\beta}_2^M \gamma_{21} \]

Total effect = Direct effect + indirect effect
Accounting for the total effects in the Gore thermometer example

<table>
<thead>
<tr>
<th>Effect</th>
<th>Total</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinton</td>
<td>0.62</td>
<td>0.51</td>
<td>0.11</td>
</tr>
<tr>
<td>Party</td>
<td>15.7</td>
<td>5.8</td>
<td>9.9</td>
</tr>
</tbody>
</table>
Other approaches to addressing confounding effects?

- Experiments
- Difference-in-differences designs
- Others?

- Is regression the best approach to addressing confounding effects?
  - Problems
Drinking and Greek Life Example

Why is there a correlation between living in a fraternity/sorority house and drinking?

- Greek organizations often emphasize social gatherings that have alcohol. The effect is being in the Greek organization itself, not the house.
- There’s something about the House environment itself.
Dependent variable: Times Drinking in Past 30 Days

C8. When did you last have a drink (that is more than just a few slips)?
- I have never had a drink \(\rightarrow\) Skip to C22 (page 10)
- Not in the past year \(\rightarrow\) Skip to C22 (page 10)
- More than 30 days ago, but in the past year \(\rightarrow\) Skip to C17 (page 8)
- More than a week ago, but in the past 30 days \(\rightarrow\) Go to C9
- Within the last week \(\rightarrow\) Go to C9

C9. On how many occasions have you had a drink of alcohol in the past 30 days? (Choose one answer.)
- Did not drink in the last 30 days
- 1 to 2 occasions
- 3 to 5 occasions
- 6 to 9 occasions
- 10 to 19 occasions
- 20 to 39 occasions
- 40 or more occasions
(14138 observations read)

. recode timespast30 timeshs (1=0) (2=1.5) (3=4) (4=7.5) (5=14.5) (6=29.5) (7=45)
timespast30: 6571 changes made
timeshs: 10272 changes made

. replace timespast30=0 if screen<=3
(4631 real changes made)
. tab timespast30

<table>
<thead>
<tr>
<th>timespast30</th>
<th>Freq.</th>
<th>Percent</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4,652</td>
<td>33.37</td>
<td>33.37</td>
</tr>
<tr>
<td>1.5</td>
<td>2,737</td>
<td>19.64</td>
<td>53.01</td>
</tr>
<tr>
<td>4</td>
<td>2,653</td>
<td>19.03</td>
<td>72.04</td>
</tr>
<tr>
<td>7.5</td>
<td>1,854</td>
<td>13.30</td>
<td>85.34</td>
</tr>
<tr>
<td>14.5</td>
<td>1,648</td>
<td>11.82</td>
<td>97.17</td>
</tr>
<tr>
<td>29.5</td>
<td>350</td>
<td>2.51</td>
<td>99.68</td>
</tr>
<tr>
<td>45</td>
<td>45</td>
<td>0.32</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>13,939</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>
Three Regressions

<table>
<thead>
<tr>
<th>Dependent variable: number of times drinking in past 30 days</th>
<th>Live in frat/sor house</th>
<th>Member of frat/sor</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.44 (0.35)</td>
<td>---</td>
<td>2.88 (0.16)</td>
<td>4.54 (0.56)</td>
</tr>
<tr>
<td>2.26 (0.38)</td>
<td></td>
<td>2.44 (0.18)</td>
<td>4.27 (0.059)</td>
</tr>
<tr>
<td>Note: Corr. Between living in frat/sor house and being a member of a Greek organization is .42</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Picture

Living in frat house

0.19

Member of fraternity

2.26

Drinks per 30 day period

2.44
Accounting for the effects of frat house living and Greek membership on drinking

<table>
<thead>
<tr>
<th>Effect</th>
<th>Total</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Member of Greek org.</td>
<td>2.88</td>
<td>2.44</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(85%)</td>
<td>(15%)</td>
</tr>
<tr>
<td>Live in frat/sor. house</td>
<td>4.44</td>
<td>2.26</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(51%)</td>
<td>(49%)</td>
</tr>
</tbody>
</table>