Addressing Alternative Explanations: Multiple Regression

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

- 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
  - Also called multiple regression, Ordinary Least Squares (OLS)
  - Presentation below is intuitive

Democratic picture

Independent picture
Republican picture

Combined data picture

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

\[ Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \epsilon_i \]

**STATATA:**

```
reg y x1 x2
reg gore clinton party3
```

**Multivariate slope coefficients**

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

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

**The Matrix form**

\[
\beta = (X'X)^{-1} X'y
\]

**The slope coefficients more simply**

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

\( X_1 \) is Clinton thermometer, \( X_2 \) is PID, and \( Y \) is Gore thermometer.
The Output

Regression analysis

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

| Variable | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|---|-----|----------------------|
| gore     | 0.51229 | 0.0175952 | 29.12 | 0.000 | 0.4777776 - 0.5467975 |
| clinton  | 5.770523 | 0.5594846 | 10.31 | 0.000 | 4.673191 - 6.867856  |
| party3   | 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 0.51 increase in the Gore thermometer.

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 linear nonrandom selection on party ID.
  - Nonrandom selection into the treatment could occur from
    - Variables other than party ID, or
    - Reverse causation, that is, feelings about Gore influencing feelings about Clinton.
  - Additionally, the regression analysis may not have entirely ruled out nonrandom selection even on party ID because it may have assumed the wrong functional form.
    - E.g., what if nonrandom selection on strong Republican/strong Democrat, but not on weak partisans

Other approaches to addressing confounding effects?

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

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 party3, cov
(observations=1745)
```

```
<table>
<thead>
<tr>
<th>Variable</th>
<th>gore</th>
<th>clinton</th>
<th>party3</th>
</tr>
</thead>
<tbody>
<tr>
<td>gore</td>
<td>660.681</td>
<td>549.991</td>
<td>13.708</td>
</tr>
<tr>
<td>clinton</td>
<td>549.991</td>
<td>881.182</td>
<td>16.905</td>
</tr>
<tr>
<td>party3</td>
<td>13.708</td>
<td>16.905</td>
<td>.8735</td>
</tr>
</tbody>
</table>
```
The Calculations

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

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

\[
= \frac{549.993 - 5.7705}{883.182} - \frac{182.883}{883.182}
\]

\[
= 0.6227 - 0.01105
\]

\[
= 0.5122
\]

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

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key explanatory variables

- Live in fraternity/sorority house
  - Indicator variable (dummy variable)
    - Coded 1 if live in, 0 otherwise
- Member of fraternity/sorority
  - Indicator variable (dummy variable)
    - Coded 1 if member, 0 otherwise
Three Regressions

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: number of times drinking in past 30 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live in frat/sor house</td>
<td>4.44 (0.35)</td>
</tr>
<tr>
<td>(indicator variable)</td>
<td>--- 2.26 (0.38)</td>
</tr>
<tr>
<td>Member of frat/sor (indicator variable)</td>
<td>--- 2.88 (0.16) 2.44 (0.18)</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.54 (0.56) 4.27 (0.059) 4.27 (0.059)</td>
</tr>
<tr>
<td>N</td>
<td>13,876 13,876 13,876</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. Corr. Between living in frat/sor house and being a member of a Greek organization is .42

The Picture

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

<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>(85%)</td>
<td></td>
<td>(15%)</td>
<td></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>(51%)</td>
<td></td>
<td>(49%)</td>
<td></td>
</tr>
</tbody>
</table>