Addressing Alternative Explanations: Multiple Regression

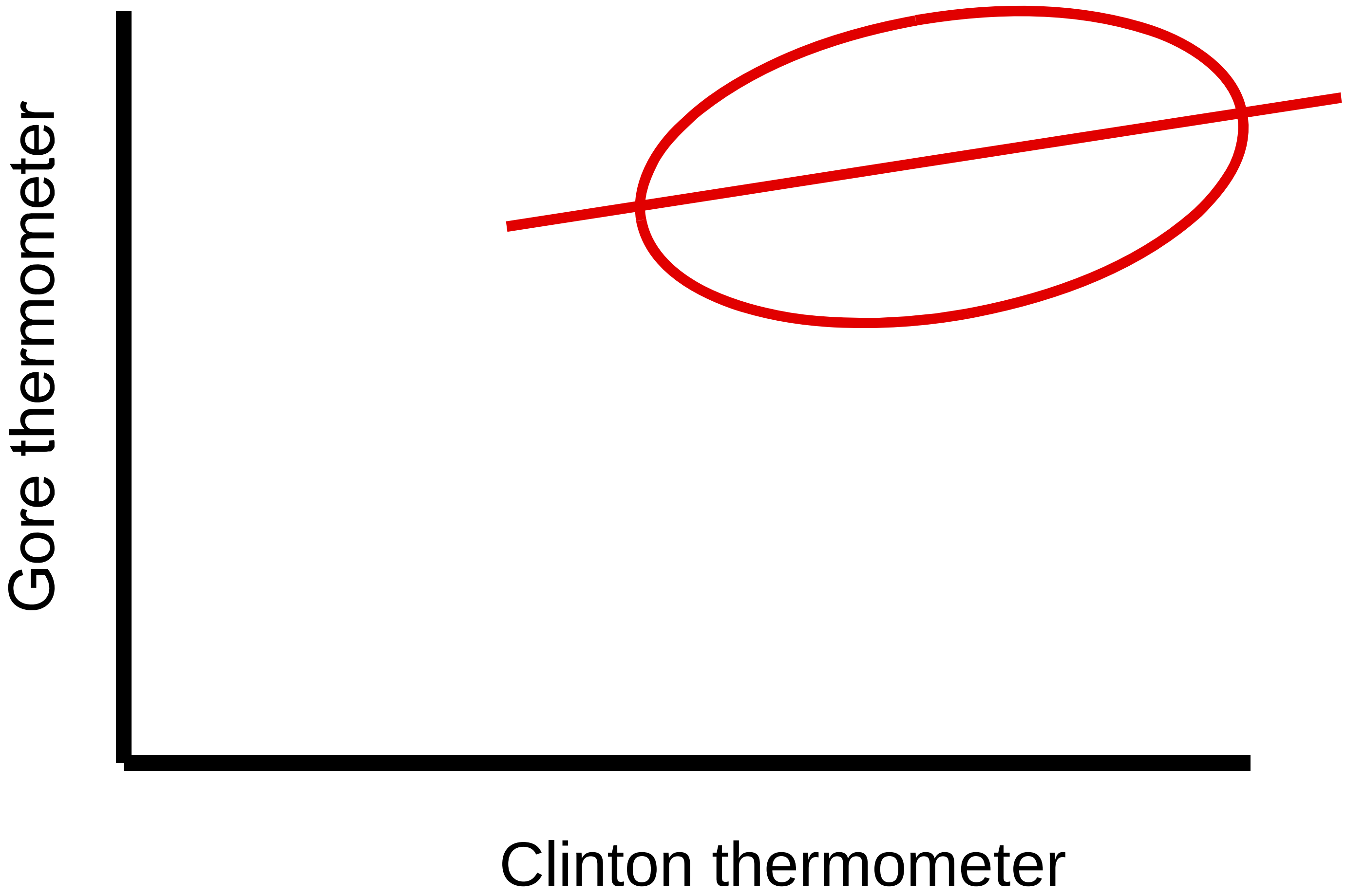
17.871

Gore Likeability Example

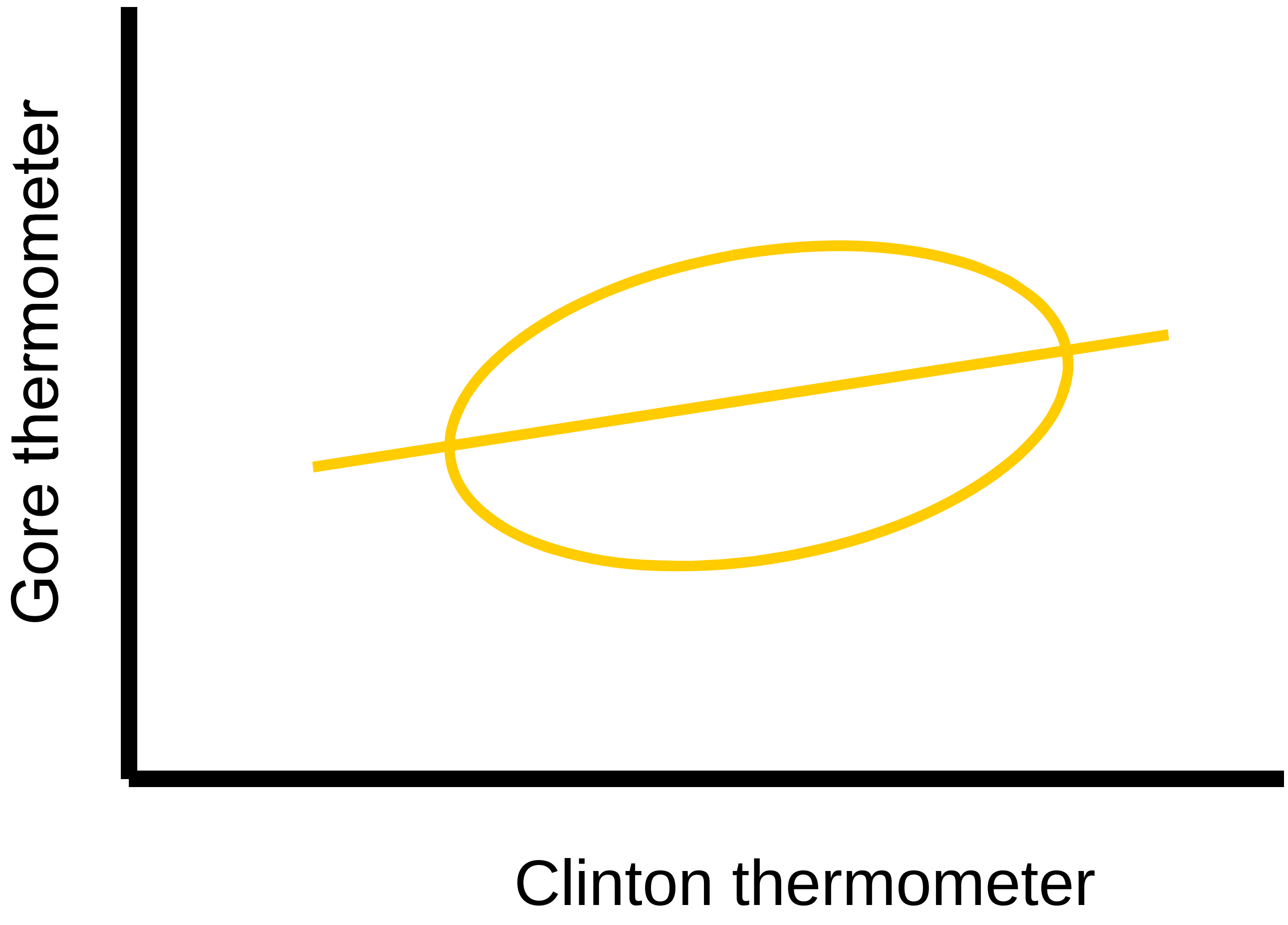
- Did Clinton hurt Gore in the 2000 election?
- How would you test this?
- What alternative explanations would you need to address?

- Other examples of alternative explanations based on omitted variables?

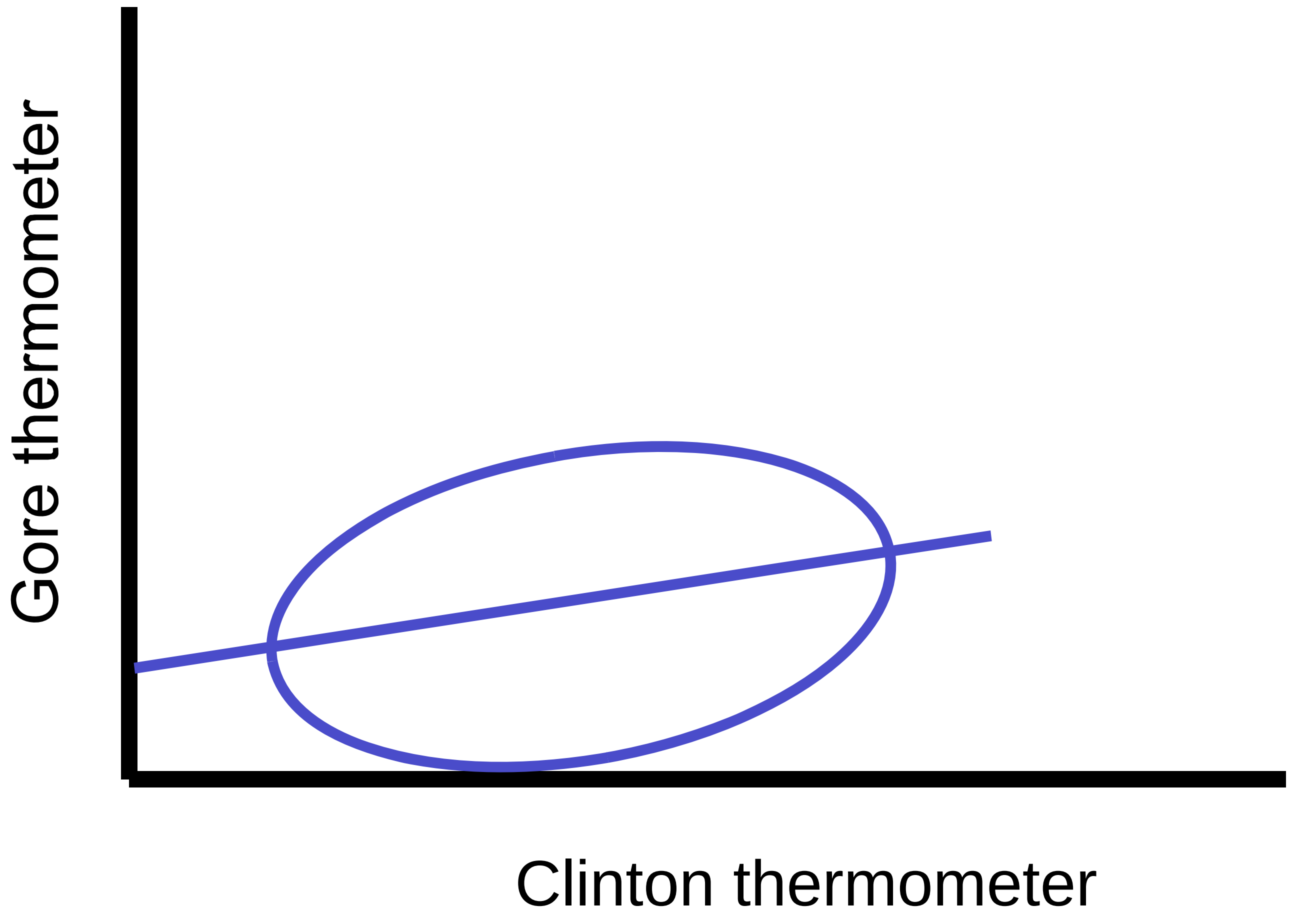
Democratic picture



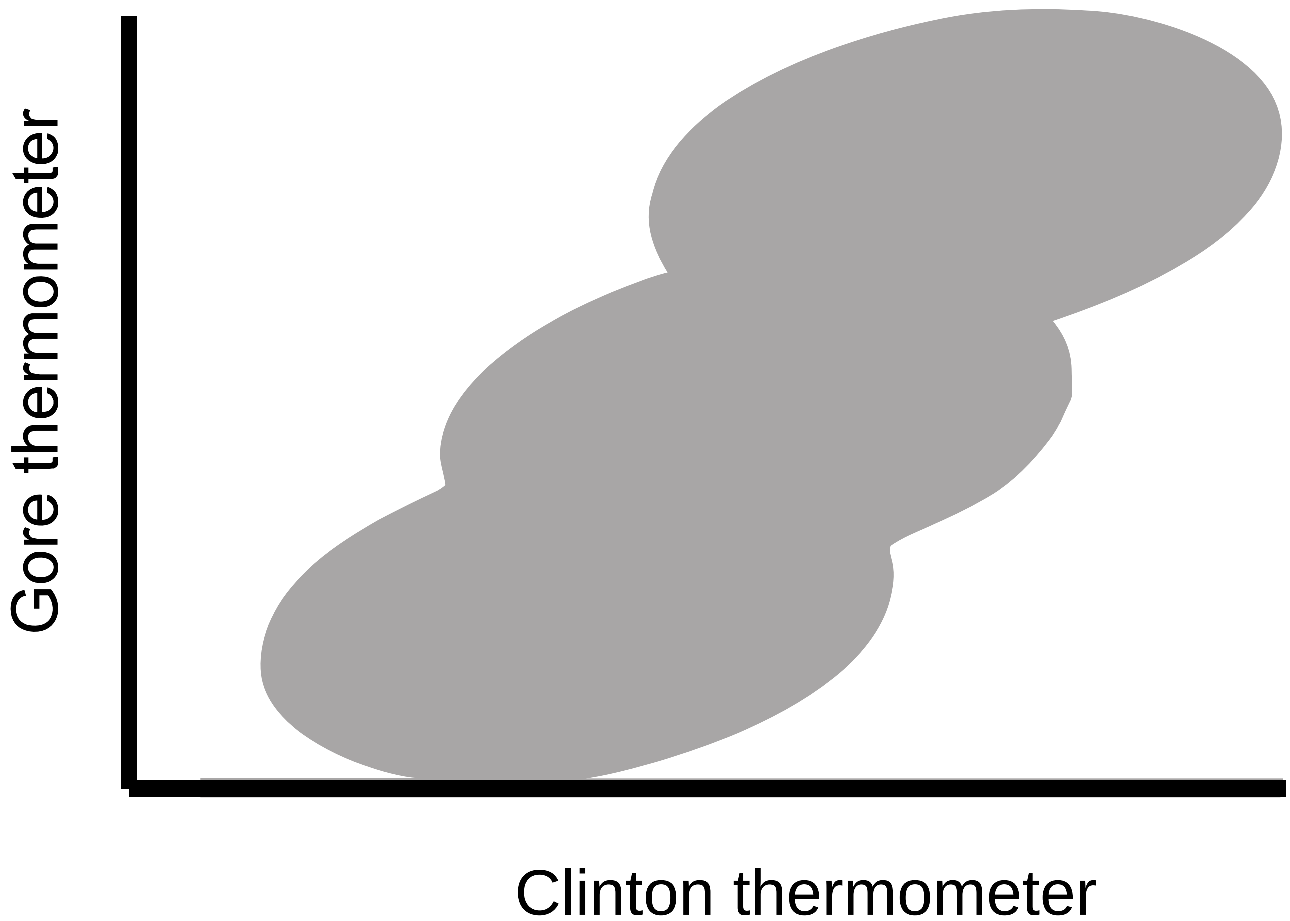
Independent picture



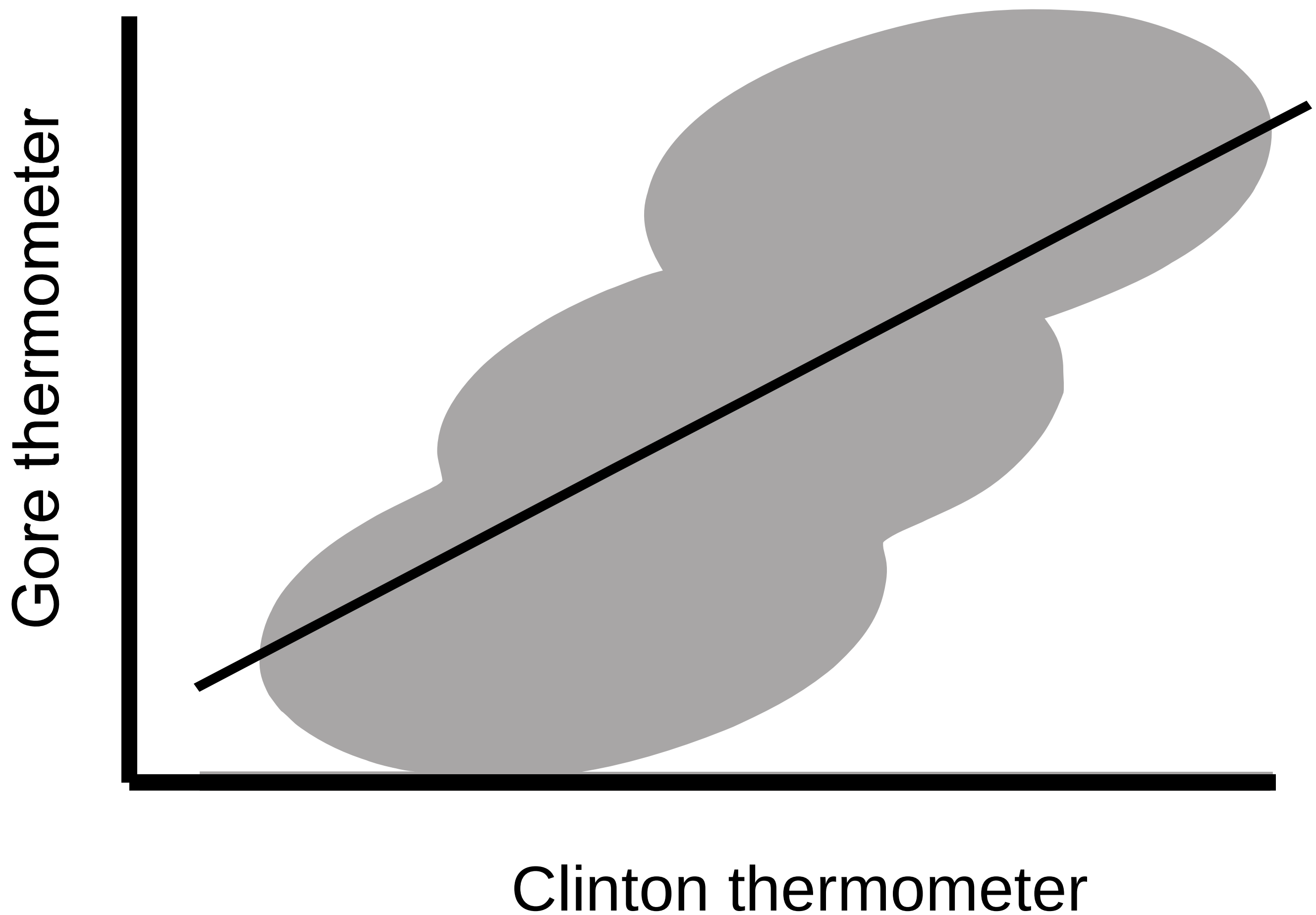
Republican picture



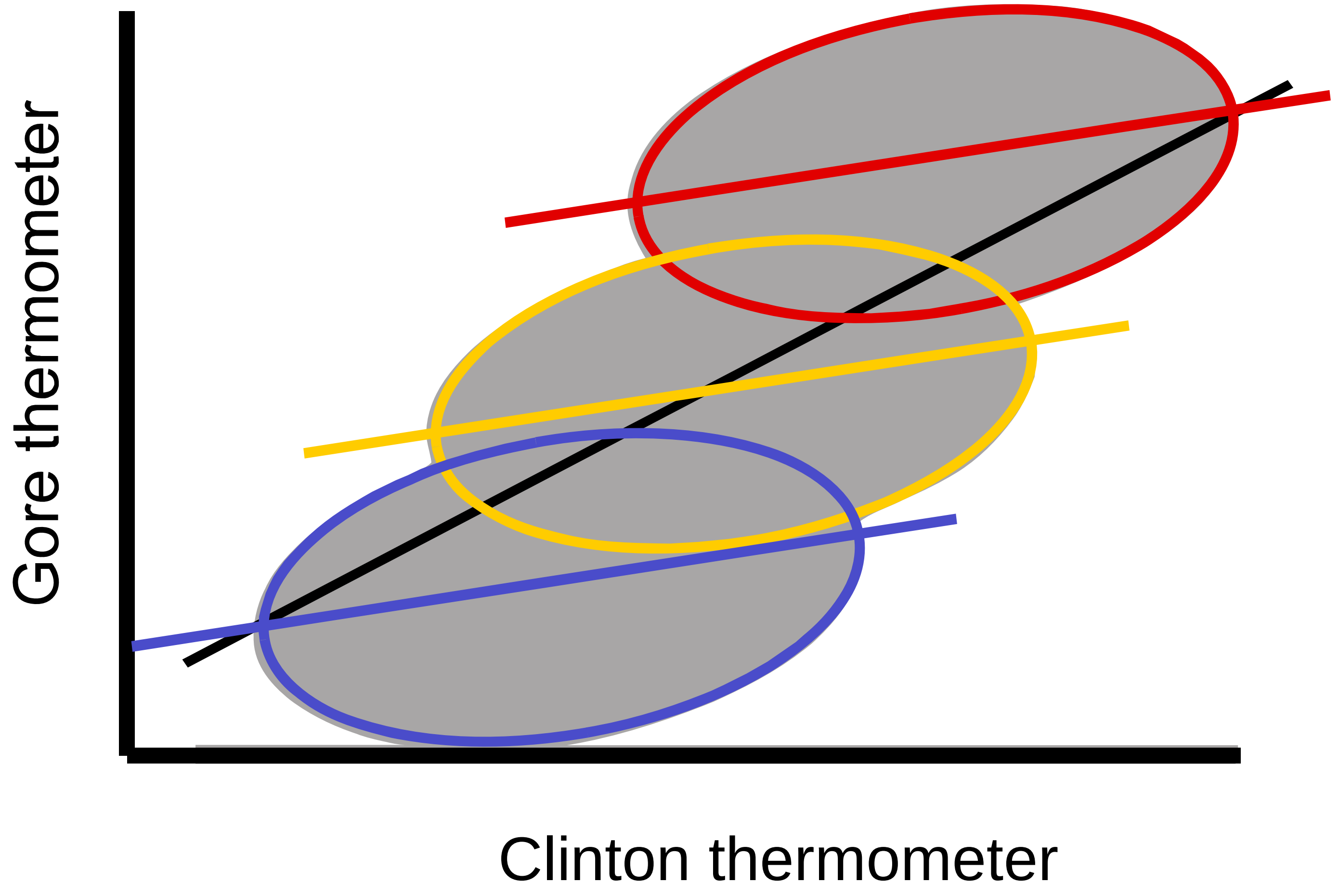
Combined data picture



Combined data picture with regression

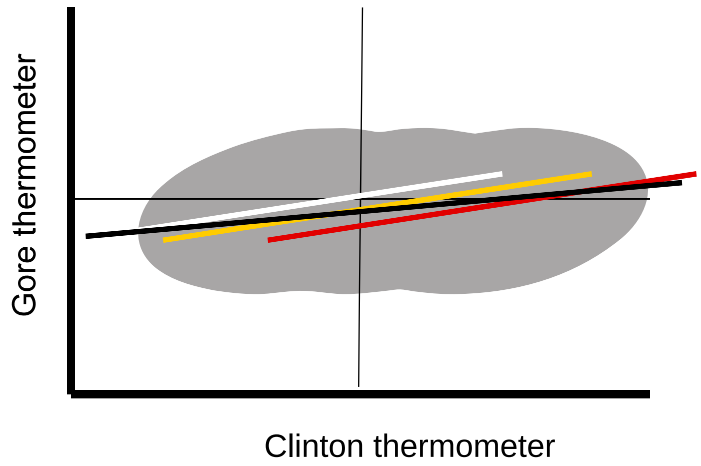


Combined data picture with "true" regression lines overlaid

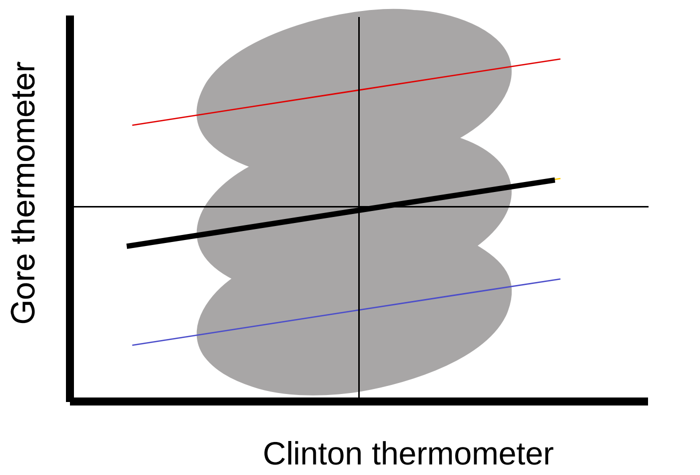


Tempting yet wrong normalizations

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



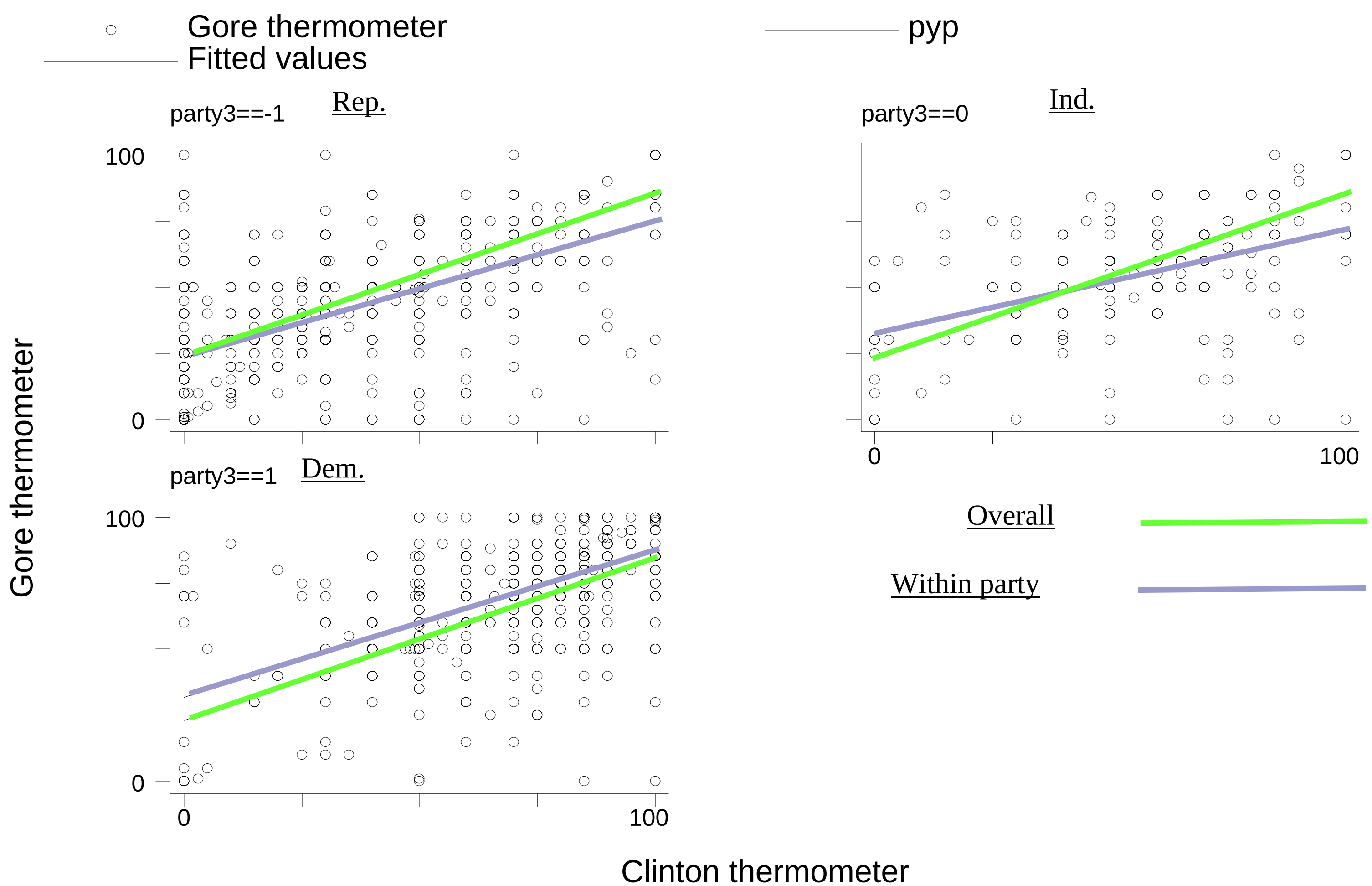
Subtract the Clinton therm. from the avg. Clinton therm. score



Summary: Why we control

- Address alternative explanations by removing confounding effects
- Improve efficiency

Gore vs. Clinton

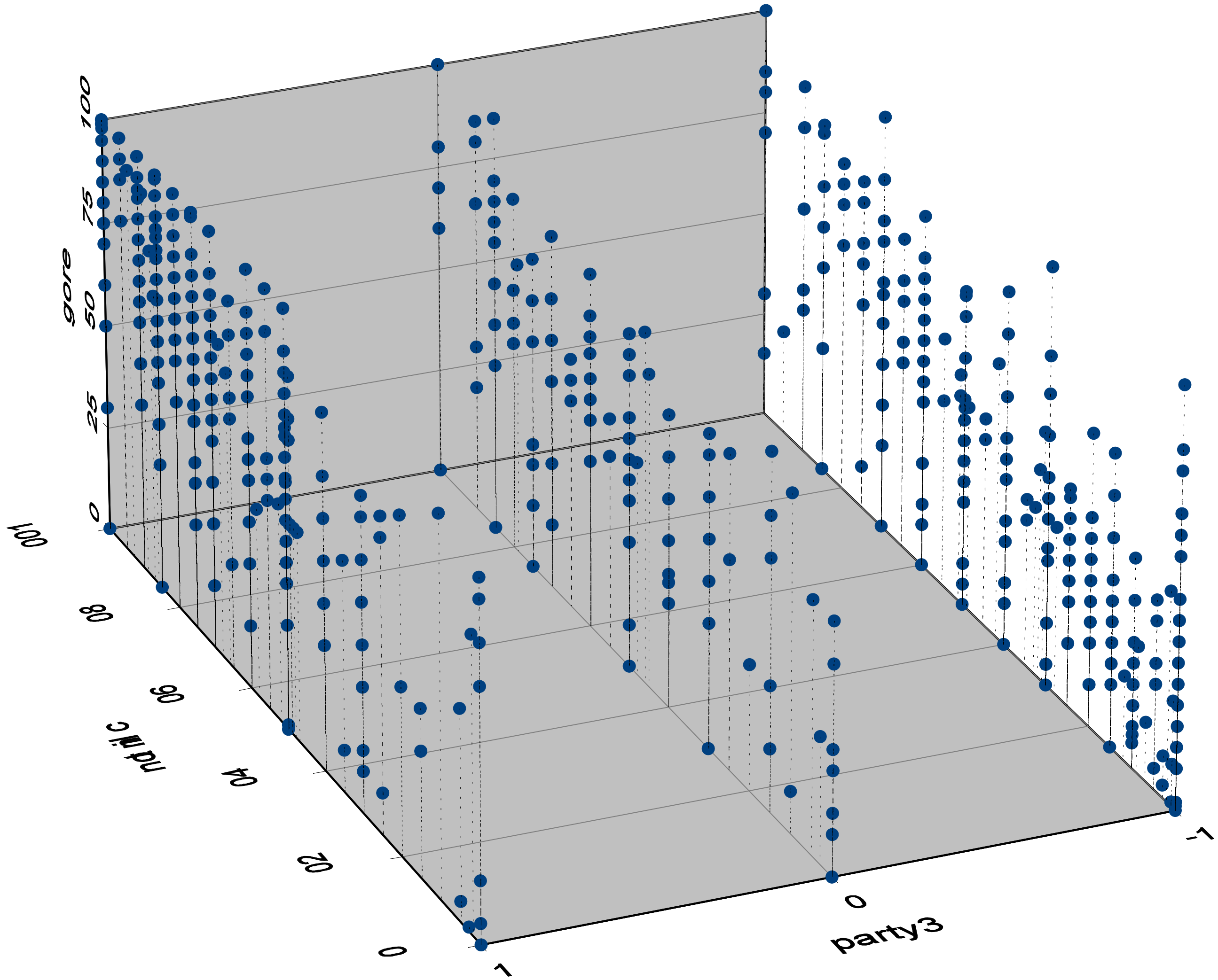


Graphs by Party (3 point scale)

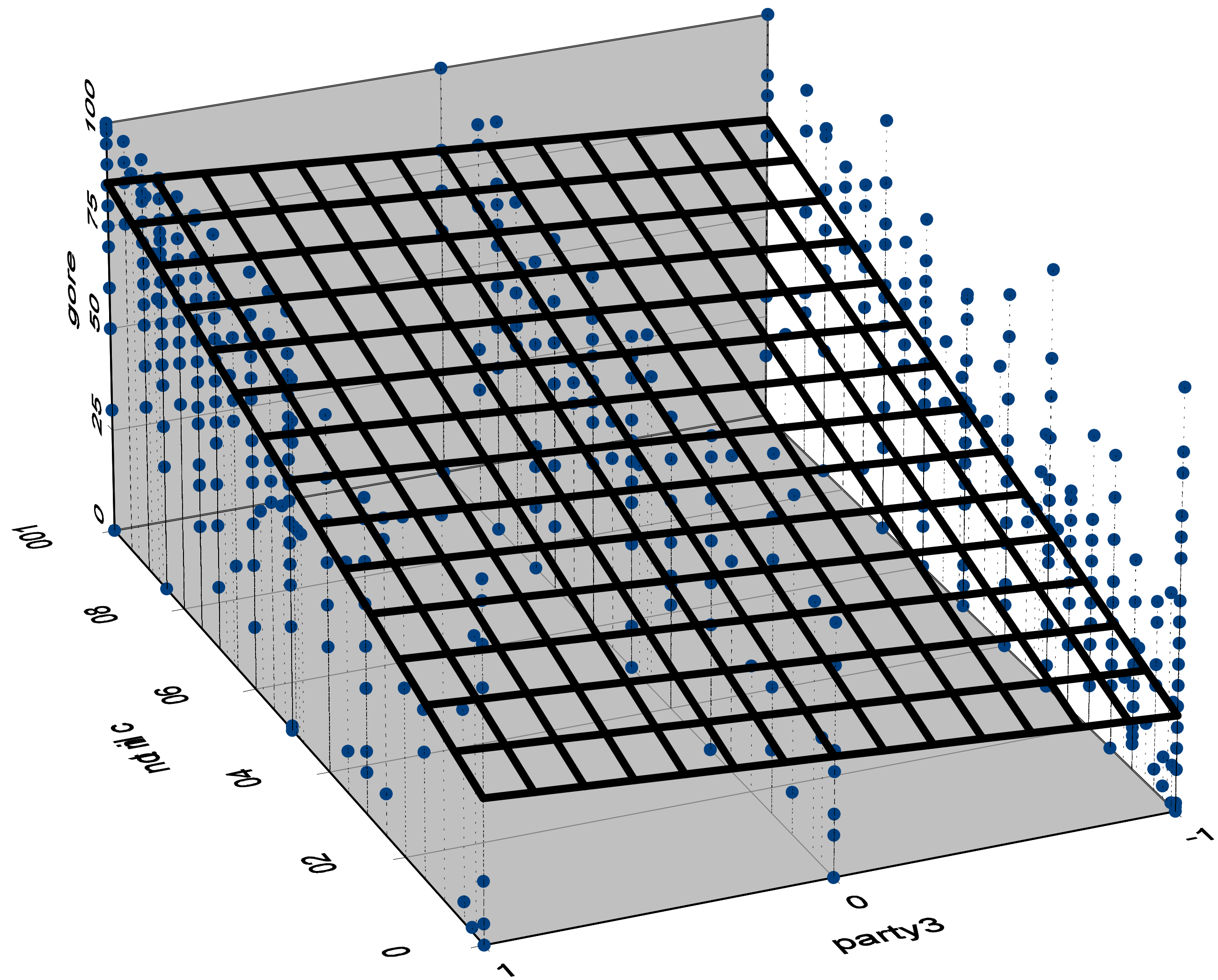
The Linear Relationship between Three Variables

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \varepsilon_i$$

3D Relationship



3D Linear Relationship



The Slope Coefficients

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

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

The Slope Coefficients More Simply

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

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

The Matrix form

y_1		1	$x_{1,1}$	$x_{2,1}$	…	$x_{k,1}$
y_2		1	$x_{1,2}$	$x_{2,2}$	…	$x_{k,2}$
…		1	…	…	…	…
y_n		1	$x_{1,n}$	$x_{2,n}$	…	$x_{k,n}$

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

Consider two regression coefficients

$$\hat{\beta}_1^B = \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)} \text{ vs.}$$

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

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

Separate regressions

	(1)	(2)	(3)
Intercept	23.1	55.9	28.6
Clinton	0.62	--	0.51
Party	--	15.7	5.8

Why did the Clinton Coefficient change from 0.62 to 0.51

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

             |     gore  clinton   party3
-------------+---------------------------
        gore |  660.681
     clinton |  549.993  883.182
      party3 |  13.7008   16.905    .8735
```

The Calculations

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

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

$$= \frac{549.993}{883.182} - 5.7705 \frac{16.905}{883.182}$$

$$= 0.6227 - 0.1105$$

$$= 0.5122$$

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

             |     gore   clinton    party3
-------------+---------------------------
        gore |  660.681
     clinton |  549.993   883.182
      party3 |  13.7008    16.905     .8735
```

The Output

```
. reg gore clinton party3

      Source |       SS       df       MS              Number of obs =    1745
-------------+------------------------------           F(  2,  1742) = 1048.04
       Model |  629261.91     2  314630.955            Prob > F      =  0.0000
    Residual | 522964.934  1742  300.209492            R-squared     =  0.5461
-------------+------------------------------           Adj R-squared =  0.5456
       Total | 1152226.84  1744   660.68053            Root MSE      =  17.327

------------------------------------------------------------------------------
        gore |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     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
------------------------------------------------------------------------------
```

Accounting for total effects

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

(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

$$\begin{tikzcd}
X_1 \arrow[dr, "\hat{\beta}_1^M"] \arrow[dd, "\gamma_{21}"'] & \\
& Y \\
X_2 \arrow[ur, "\hat{\beta}_2^M"'] &
\end{tikzcd}$$

Accounting for the total effects in the Gore thermometer example

Effect	Total	Direct	Indirect
Clinton	0.62	0.51	0.11
Party	15.7	5.8	9.9

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 sips)?

- ◯ I have never had a drink ⟶ Skip to **C22** (page 10)
- ◯ Not in the past year ⟶ Skip to **C22** (page 10)
- ◯ More than 30 days ago, but in the past year ⟶ Skip to **C17** (page 8)
- ◯ More than a week ago, but in the past 30 days ⟶ Go to **C9**
- ◯ Within the last week ⟶ Go to **C9**

C9. On how many occasions have you had a drink of alcohol in the <u>past 30 days</u>? (Choose one answer.)

- 1 ◯ Did not drink in the last 30 days
- 2 ◯ 1 to 2 occasions
- 3 ◯ 3 to 5 occasions
- 4 ◯ 6 to 9 occasions
- 5 ◯ 10 to 19 occasions
- 6 ◯ 20 to 39 occasions
- 7 ◯ 40 or more occasions

```
. infix age 10-11 residence 16 greek 24 screen 102
timespast30 103 howmuchpast30 104 gpa 278-279 studying 281
timeshs 325 howmuchhs 326 socializing 283 stwgt_99 475-493
weight99 494-512 using da3818.dat,clear
(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)
```

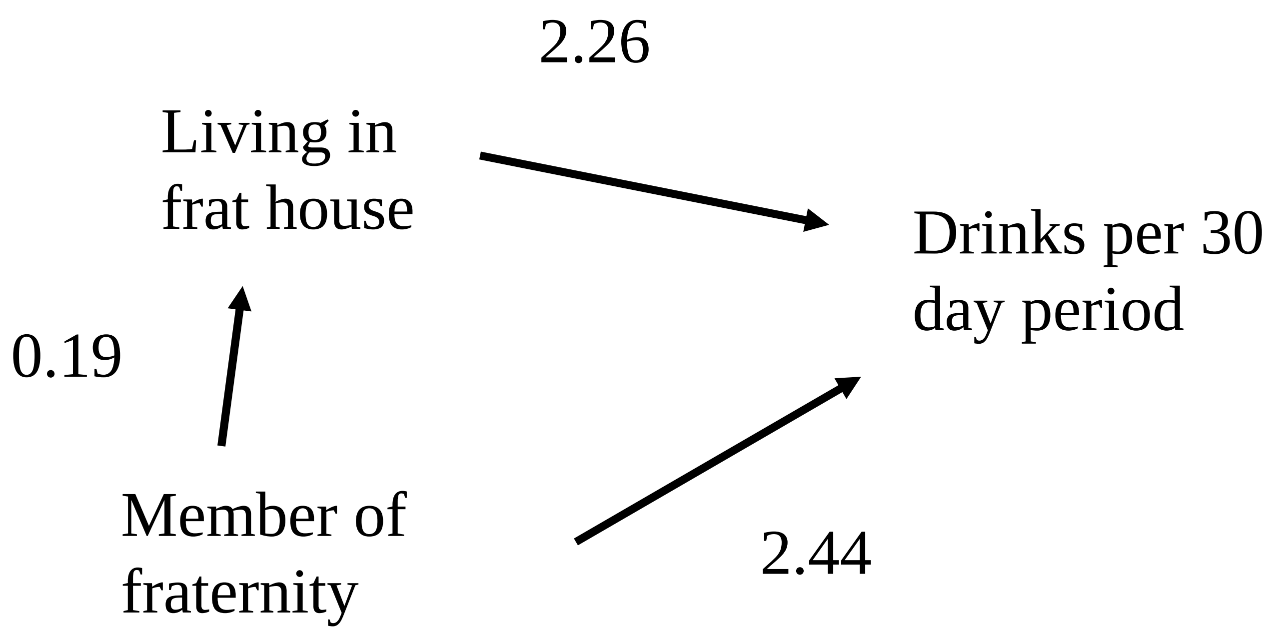
```
. tab timespast30

timespast30 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      4,652       33.37       33.37
        1.5 |      2,737       19.64       53.01
          4 |      2,653       19.03       72.04
        7.5 |      1,854       13.30       85.34
       14.5 |      1,648       11.82       97.17
       29.5 |        350        2.51       99.68
         45 |         45        0.32      100.00
------------+-----------------------------------
      Total |     13,939      100.00
```

Three Regressions

Dependent variable: number of times drinking in past 30 days			
Live in frat/sor house	4.44 (0.35)	---	2.26 (0.38)
Member of frat/sor	---	2.88 (0.16)	2.44 (0.18)
Intercept	4.54 (0.56)	4.27 (0.059)	4.27 (0.059)
R2	.011	.023	.025
N	13,876	13,876	13,876

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

The Picture


2.26
Living in
frat house
Drinks per 30
day period
0.19
Member of
fraternity
2.44

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

Effect	Total	Direct	Indirect
Member of Greek org.	2.88	2.44 (85%)	0.44 (15%)
Live in frat/ sor. house	4.44	2.26 (51%)	2.18 (49%)