

From previous lecture

$$\log_{10}(\widehat{\text{Total comp}}) = 2.48 + 0.43 \log_{10}(\text{net sales}) + 0.082 \text{ Energy} + \dots$$

$$\log_{10}(\widehat{\text{Total comp}})_t = 2.48 + 0.43 \log_{10}(\text{net sales})_t$$

for CEOs in
discretionary
ind.

$$\log_{10}(\widehat{\text{Total comp}})_t = 2.48 + 0.43 \log_{10}(\text{net sales})_t + 0.082$$

for CEOs
in energy
industry

⋮
Intercept

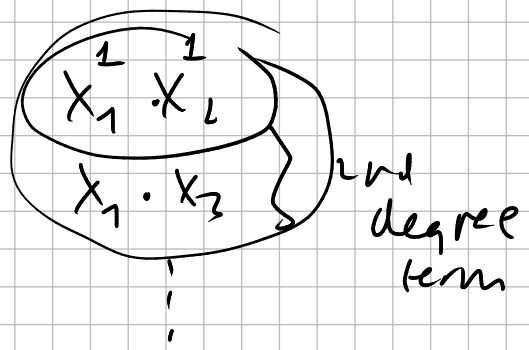
Why is "regardless of industry" incorrect?

$$\log_{10}(\widehat{\text{comp dis}}) = 2.48 + 0.43 \log_{10}(\text{net sales})_{\text{old}}$$

$$\log_{10}(\widehat{\text{comp energy}}) = 2.48 + 0.082 + 0.43 \log_{10}(\text{net sales})_{\text{new}}$$

When we compare firms with the same level of net sales, CEOs in energy industry on average earn about 8.2% more than CEOs in discretionary industry.

Y regressand
 X_1^1
 X_2^1
 X_3^1 regressors



\rightarrow in R form
 $\text{lm}(Y \sim X_1 + X_2 + X_1:X_2)$
 interaction term $X_1 * X_2 \rightarrow$ math form
 main terms

$\text{lm}(Y \sim X_1 + X_2 + X_3 + X_1:X_2 + X_1:X_3 + X_2:X_3)$

also possible to have interaction terms of the form $X_1 * X_2 * X_3$
 \rightarrow very rare

difference in difference approaches to causal inference (DiD)

$$\log_{10}(\widehat{\text{Total Comp Mil}})_t = 2.05 + 0.47 * \log_{10}(\text{net sales})$$

$$+ 0.487 * \text{Energy} + \dots - 1.21 * \text{Utility}$$

$$- 0.043 * \log_{10}(\text{net sales}) * \text{Energy} + \dots$$

Discretionary industry: $\log_{10}(\widehat{\text{Total Comp Mil}})_t$

$$= 2.05 + 0.47 * \log_{10}(\text{Net sales})$$

Energy industry: $\log_{10}(\widehat{\text{Total Comp Mil}})_t = 2.05 + 0.47 * \log_{10}(\text{net sales})$

$$+ 0.487 - 0.043 * \log_{10}(\text{net sales})$$

$$= 2.537 + 0.427 * \log_{10}(\text{net sales})$$

ATTEND.RAW is an example of tab-delimited data

$$100 \text{ percent} - 99 \text{ percent} = 1 \text{ percentage point}$$

vs 1 percentage point
1 percent

$\left. \begin{array}{l} 10 \text{ percent} \\ 20 \text{ percent} \end{array} \right\} 10 = 20 - 10$
 or 100% (relative diff)

$$= \frac{20 - 10}{10}$$

-0.001156

who

When we compare students have the same prior GPA but whose attendance rate differ by 1 percentage point, the student with the higher attendance rate on average has 0.0012 lower standardized final exam score compared to the student with the lower attendance rate.

Intercept:

The average standardized final exam score for students whose

attendance rate is 75 percent and prior GPA is 2.6
 (c. attend rate = 0 @ attend rate = 75) (c. pri GPA = 0 @ pri GPA = 2.6)

is -0.045 . This is
 \downarrow
 standardized scale

$-0.045 \times (\underbrace{\text{sd}(\text{attend} \& \text{final})}_{4.7}) \approx -0.21$
 points

in terms of the original final exam scores.

$$\widehat{\text{stdnt}}_t = 0.923 \text{ atndrte}_t - 0.03 \text{ atndrte}_t^2 - 0.47 \text{ pri6PA}_t + 0.013 \text{ atndrte}_t \times \text{pri6PA}_t$$

All students

"linear" in linear regression

$$\widehat{Y}_t = X_t' \beta$$

regression line/surface

$$X_t' = (1, \log X_{1t}, X_{2t}, X_{2t}^2, \dots)$$