From previous lecture

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

$$\widehat{\log_{10}(\text{Total comp})}_t = 2.48 + 0.43 \log_{10}(\text{net sales})_t$$ for CEOs in discretionary ind.

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

⋮

intercept

Why is "regardless of industry" incorrect?

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

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

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

$X_1^1 \cdot X_2^1$

$X_1 \cdot X_3$

2nd degree term

$\rightarrow$ in R form

lm(Y ~ X1 + X2 + X1:X2)

main terms

interaction term

$X_1 * X_2$ $\rightarrow$ math form

lm(Y ~ X1 + X2 + X3 + X1:X2 + X1:X3 + X2:X3)

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)

$$\widehat{\log_{10}(\text{TotalCompMil})}_t = 2.05 + 0.47 * \log_{10}(\text{net sales})$$
$$+ 0.487 * \text{Energy} + \dots - 1.21 * \text{Utility}$$
$$- 0.043 * \log_{10}(\text{net sales}) * \text{Energy} + \dots$$

Discretionary industry: $\widehat{\log_{10}(\text{TotalCompMil})}_t$
$$= 2.05 + 0.47 \log_{10}(\text{Netsales})$$

Energy industry: $\widehat{\log_{10}(\text{TotalCompMil})}_t = 2.05 + 0.47 \log_{10}(\text{netsales})$
$$+ 0.487 - 0.043 * \log_{10}(\text{net sales})$$
$$= 2.537 + 0.427 * \log_{10}(\text{net sales})$$

ATTEND.RAW is an example of tab-delimited data

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

vs 1 percentage point
1 percent

10 percent
20 percent } 10 percentage points (absolute diff) $= 20 - 10$
or
100% (relative diff)
$= \frac{20-10}{10}$

-0.001156

When we compare students who 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 stdized final exam score for students whose attendance rate is 75 percent and prior GPA is 2.6
(c.atndrte = 0 ⟺ atndrte = 75) (c.priGPA = 0 ⟺ priGPA = 2.6)
is $\underbrace{-0.045}_{\text{standardized scale}}$. This is $-0.045 * (\underbrace{sd(attend\$final)}_{4.7}) \approx -0.21$ points

in terms of the original final exam scores.

$$\widehat{stndfnl}_t = 0.923 - 0.03\, atndrte_t - 0.47\, priGPA_t + 0.013\, atndrte_t \cdot priGPA_t$$

all students

"linear" in linear regression

$$\hat{Y}_t = X_t' \boxed{\hat{\beta}}$$ regression line / surface

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