Returns to Schooling

- MM Section 6.1 gives another example of regression and OVB in observational data.
  - Read pages 209–211.
- This regression includes a quadratic term.
  - The dependent variable is earnings. The main covariate is experience.
  - It includes both a linear experience term and an experience-squared term to show that earnings increase with experience but increase more slowly in later years.

Experience as a Covariate

- The goal is to estimate the returns to schooling.
- Mincer included work experience as a covariate because people with less schooling but much more work experience often earn more than people with more schooling but no work experience.
- Note in Equation 6.2 that the short version of the regression gives a coefficient of 0.07 on years of schooling, while the long version gives a coefficient of .107.

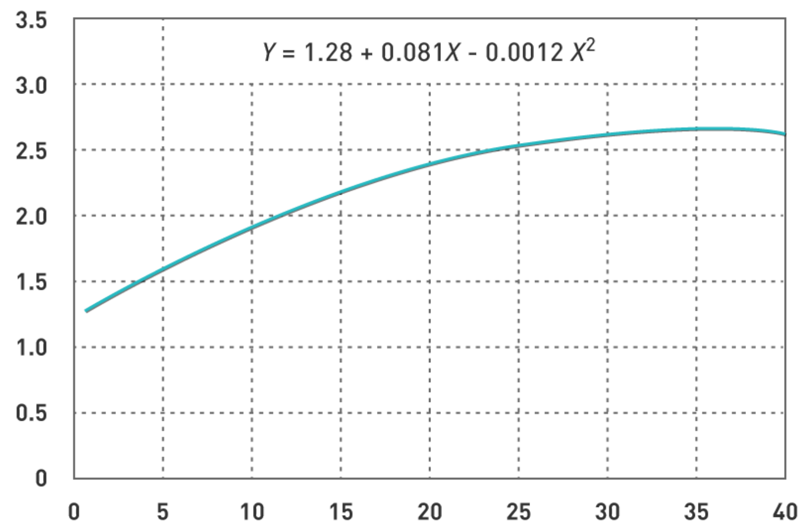
$$\ln Y_i = \alpha + \underset{(.002)}{.070}\, S_i + e_i$$

$$\ln Y_i = \alpha + \underset{(.001)}{.107}\, S_i + \underset{(.001)}{.081}\, X_i - \underset{(.00002)}{.0012}\, X_i^2 + e_i$$

- Why does this happen?

Experience as a Covariate (cont.)

- Omitted variable bias (OVB) causes us to underestimate the returns to schooling. Why?
  - This is observational data, so the two regressors can be correlated with each other. Here, schooling and work experience are negatively correlated.
  - Both schooling and experience have positive effects on earnings. When we omit experience, we force the effect of experience on earnings to become part of the schooling coefficient.
  - As people with more schooling have less experience, when we increase schooling, earnings don't increase as much as they would if we were holding experience constant as a covariate.
  - Hence, with OVB we measure the coefficient of interest to be 7% instead of 11%.

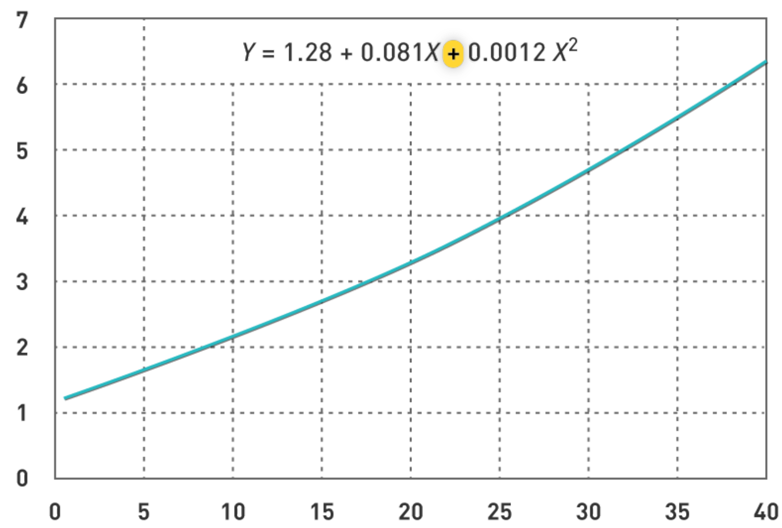


Quadratic Specification: Equation 6.2

- The quadratic specification in experience allows for the benefits of experience to accrue at a declining rate. (Note that we use years of potential experience as a proxy for actual years of work experience, since we can't always measure the latter.)

$$\ln Y_i = \alpha + \underset{(.001)}{.107}\, S_i + \underset{(.001)}{.081}\, X_i - \underset{(.00002)}{.0012}\, X_i^2 + e_i$$

- A positive coefficient on the linear term and a negative (smaller) coefficient on the squared term: concave function.
- A positive coefficient on the linear term and a positive coefficient on the squared term: convex function.



Reading

Read MM from the bottom of page 211 to the middle of page 214.

"Is control for experience sufficient for *ceteris* to be *paribus*?"

Ability as an Omitted Variable

- Griliches expected Mincer's estimates to be overstated due to the omission of ability from the regression. Years of schooling are generally positively correlated with ability.
- When Griliches included IQ as a covariate, the estimated returns to schooling fell from 6.8% per year of schooling to 5.9%. Without IQ, OVB caused an overestimate of returns to schooling.
- Have we now controlled for everything that might cause biased results?
    - No. There are more kinds of ability than IQ (e.g., emotional intelligence, curiosity, etc.). We still have omitted variables that are likely to be correlated with years of schooling.
    - Thus, we cannot entirely trust the ~6% per year number.

Ability as an Omitted Variable (cont.)

- How could we be sure of the right result?
    - We would need an experiment that randomly sends some students to more years of school than others.
    - Then every possible omitted variable would be uncorrelated with years of schooling, eliminating OVB.

Attenuation Bias

- Angrist and Pischke mention a concept usually called **attenuation bias**.
- Imagine that we don't always correctly measure the treatment variable (years of schooling). With measurement error in the X variable, the resulting coefficient is biased toward zero.
- Consider my effort to measure the effects of online advertising, in which we matched Yahoo! users to retail purchases using names and e-mail addresses. Suppose the matching procedure allowed for nonexact matches.
    - I might have some purchasers who I thought were in the control group, but who were really in the treatment group (and vice versa).
    - Thus, some of the advertising effects would appear in the control group rather than in the treatment group, which means the effects would look smaller than actual.
- This is an example of attenuation bias. The X variable is treatment assignment, and it is measured with error, so the treatment effect will be biased toward zero.

Reading

- Optional reading: For more about the theory of this concept, you can read MM pages 240–241 in the Chapter 6 appendix.
- Next, read MM pages 214–217 on **bad controls**, a type of covariate we do not want in our regressions when we analyze experiments.

When you analyze experimental results, do not include other outcome variables as covariates on the right-hand side of the regression.

Example: Random College Assignment

Table 6.1 How Bad Control Creates Selection Bias

<table>
<tr><th rowspan="2">Type of worker</th><th colspan="2">Potential occupation</th><th colspan="2">Potential earnings</th><th colspan="2">Average earnings by occupation</th></tr>
<tr><th>Without college (1)</th><th>With college (2)</th><th>Without college (3)</th><th>With college (4)</th><th>Without college (5)</th><th>With college (6)</th></tr>
<tr><td>Always Blue (AB)</td><td>Blue</td><td>Blue</td><td>1,000</td><td>1,500</td><td rowspan="2">Blue
1,500</td><td>Blue
1,500</td></tr>
<tr><td>Blue White (BW)</td><td>Blue</td><td>White</td><td>2,000</td><td>2,500</td><td rowspan="2">White
3,000</td></tr>
<tr><td>Always White (AW)</td><td>White</td><td>White</td><td>3,000</td><td>3,500</td><td>White
3,000</td></tr>
</table>

- True average treatment effect (ATE) is $500.

$$Y_i = \alpha + \beta E_i + \gamma W_i + e_i$$

- A regression on both a "college education" dummy E_i and a "white-collar occupation" dummy W_i yields a coefficient of $\beta = 0$, and will mistakenly indicate that the return to college education is $0.
- This happens because only the most talented non-college-educated workers will take white-collar jobs, and only the least talented college-educated workers will take blue-collar jobs.

More About Bad Controls

- We generally want to know the total effect of schooling on earnings, because it helps you become a data scientist, and because more educated data scientists earn more than less educated data scientists.
    - Including the occupation covariate is therefore a bad idea, because it picks up only the latter kind of variation.

Table 6.1 How Bad Control Creates Selection Bias

<table>
<tr><th rowspan="2">Type of worker</th><th colspan="2">Potential occupation</th><th colspan="2">Potential earnings</th><th colspan="2">Average earnings by occupation</th></tr>
<tr><th>Without college (1)</th><th>With college (2)</th><th>Without college (3)</th><th>With college (4)</th><th>Without college (5)</th><th>With college (6)</th></tr>
<tr><td>Always Blue (AB)</td><td>Blue</td><td>Blue</td><td>1,000</td><td>1,500</td><td rowspan="2">Blue
1,500</td><td>Blue
1,500</td></tr>
<tr><td>Blue White (BW)</td><td>Blue</td><td>White</td><td>2,000</td><td>2,500</td><td rowspan="2">White
3,000</td></tr>
<tr><td>Always White (AW)</td><td>White</td><td>White</td><td>3,000</td><td>3,500</td><td>White
3,000</td></tr>
</table>

Example: eBay Reputation and Seller Revenue

- We want to know whether having a higher eBay reputation causes the seller to earn more revenue on eBay.
- So we do an experiment with two eBay seller accounts, one of which has a higher reputation score.
    - In the regression analysis, we might set:
        - Outcome Y = auction price
        - Treatment D = seller reputation
        - Covariate X = number of bids
    - But, number of bids would be a bad control.
    - General principle: It's a bad idea to include posttreatment outcomes as covariates.

Fundamentally Unanswerable Questions, or FUQs

- Some research questions are poorly posed, e.g., "What is the effect on earnings of being born in Africa instead of North America?"
- What experiment would answer this question?
    - Randomly assign some Americans to be born in Africa, and let their families move back as soon as childbirth is over?
    - Randomly assign some Americans to live their whole lives in Africa, to experience being the out-of-place, pale-skinned person in their community?
    - Randomly assign some African families to move to America permanently, and compare their kids to those randomly assigned to stay in Africa?

Fundamentally Unanswerable Questions, or FUQs (cont.)

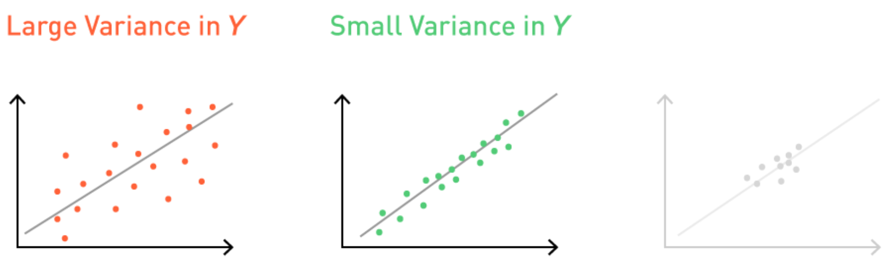
- It seems impossible to do an experiment that compares "the experience of living your whole life in Africa as a person of African descent" and "the experience of living your whole life in North America as a person of the white majority culture in the United States."
    - This question is fundamentally unanswerable, or, in a word, FUQ'd.

Questions We Should Ask in Any Causal-Effect Research

1. What is the causal relationship of interest?
2. What is the ideal experiment to measure this? (Ask this even if you're doing observational research.)
    - If your question seems FUQ'd, is there a different but related question that can be answered with a hypothetical experiment?
3. What is your identification strategy?
    - That is, where does your variation come from, and why do you think it is independent of potential outcomes?
4. How are you computing your confidence intervals?
    - This question is the focus of our next topic.

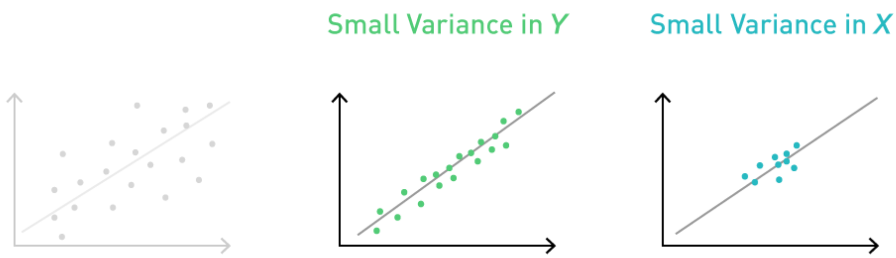
Reading

- Read MM, Chapter 2 appendix, pages 95–97.



Standard Errors and Confidence Intervals in Regression

- How precisely have we estimated our slope coefficient?
  - Figuring out where "noise" comes from
- In general, standard errors are
  - Larger when variance of Y is larger
  - Smaller when variance of X is larger



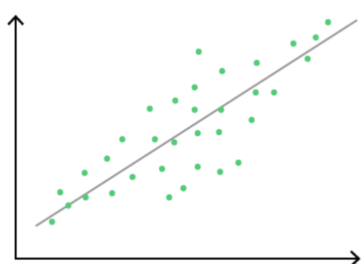
Standard Errors in Regression Output

- Treatment effect (with 95 percent confidence interval) = slope coefficient ± 2 standard errors
- OLS standard errors assume each observation's idiosyncratic component ε is independent and identically distributed
- Independence makes sense in randomized experiment
- Why do we expect same variance?

Heteroskedasticity

- Different observations with different error variances are **heteroskedastic**
- Contrast with simpler case of homoskedasticity
- Does variance of Y around regression line change as X changes?
- Homoskedasticity is the default assumption under which OLS standard errors are usually computed
- Vertical error variance causes uncertainty about line's true slope

Which Measures Slope More Accurately?



- Fanned-out data means the slope is very uncertain.
- When data is most certain at endpoints, plot is more accurate; endpoints anchor slope.
- Leverage: Data points nearer ends of regression line influence slope more.

Robust Standard Errors

- RSEs give accurate confidence intervals when error variance varies with *X*
  - Also known as heteroskedasticity-robust standard errors or Eicher-White standard errors
- No need to know shape of heteroskedasticity or which *X*variables correlate with variance
  - Optional: technical details and matrix algebra, see p. 45, *MHE*
- Very big/very small *X* values have more leverage on slope coefficient β

How Formula Accounts for Leverage

- When estimating variance of $\hat{\beta}$
  - Take weighted sum of squared residuals (i.e., squared vertical deviations from regression line)
  - Divide by total variance in *X*
  - Weights in weighted sum correspond to leverage of each observation
  - Squared residuals ($\hat{\varepsilon}$-squared) weighted by squared horizontal deviations from mean: $(X_i - \bar{X})^2$

Tennessee STAR Experiment

- Randomized at classroom level
- Each classroom's students had same teachers, similar backgrounds/experiences
- More similarity within each classroom than between classrooms
- Changing 20 similar students from control group to treatment group moves potential *Y*for all 20 at once
    - Result: more variance than if 20 randomly chosen people had been moved
- CSEs account for lack of independence, as RSEs account for heteroskedasticity
- E.g., in STAR experiment, CSEs are about three times bigger than OLS standard errors
    - Further technical details, see *MHE* 8.2

Takeaways

- Best to use always RSE (i.e., not assume homoskedasticity)
- If treatment assignment is clustered, also use CSE to avoid unintentionally overstating precision

Multifactor Experiments

- Chapter 9 of Gerber and Green's *Field Experiments*
    - 9.3.3 Multi-Dimensional Treatment
    - **Treatment-by-treatment interaction:** Running regressions with interaction terms to tell us how much more one treatment matters when the other treatment is turned on

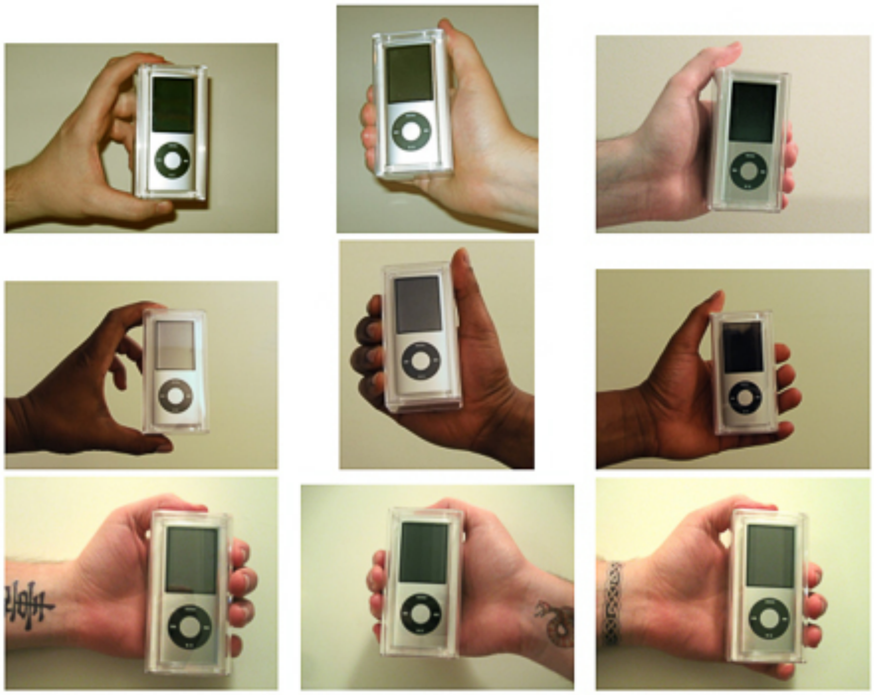
Example: Doleac and Stein (2013)

- Used Craigslist experiment to assess race- and class-based discrimination among consumers
  - Ran ads to sell iPods in different local markets
  - Measured average offer
  - Measured average number of offers
- Varied three dimensions

Jennifer L. Doleac and Luke C.D. Stein, "The Visible Hand: Race and Online Market Outcomes,"*The Economic Journal* Volume 123 Issue 572 (2013): F469-F492, doi: 10.1111/ecoj.12082.

Doleac and Stein: Three Dimensions

- Race or class
  - White hand
  - White hand with tattoo
  - Black hand
- Ad quality
  - Grammar or spelling errors in half the ads
  - No errors in the other half
- Asking price ("or best offer")
  - $90, $110, $130



Craigslist Experiment Ad Quality

	High-quality advertisement	Low-quality advertisement
Ad title	NEW 5th gen silver 8gb IPOD NANO – NEVER OPENED!	NEW 5th gen silver 8gb IPOD NANO – NEVER OPENED!
Ad text	I have a brand new iPod Nano that I don't need (I already have one). Silver, 8GB, 5th generation (with video), never opened! Holds up to 8 hours of video or 2000 songs. 24-hour battery life. This is the latest version! It sells for $149 plus tax in the store, but I'll let it go for $[price] or the best offer I receive. So, make me an offer! I'll meet you wherever is convenient. We can meet wherever is convenient for you.	ive got a brand new ipod nano that i dont need (already have one). silver, 8GB, 5th generation (with video), never opened at all!! holds up to 8 hours of video or 2000 songs. 24-hour batttery life. this is the newestt version!! it sells for $149 plus tax in teh store, but i'll let it go for $[price] or the best offer i get. so make me an offer!!! i'll meet you wherevers good for you.

Doleac and Stein: Results

- Both race and tattoo had much more treatment effect than ad quality.

Treatment-by-Treatment Interactions

- Consider the question, "Do Black sellers hurt themselves more with bad grammar than White sellers do?"
  - Read 9.3.3 with this in mind.
  - Ignore the last two paragraphs containing subtle points.

2x2 Design

- Four treatments summarized with a 2x2 table

	Colin	Jose
Good grammar	52%	37%
Bad grammar	29%	34%

Craigslist Experiment Design

- 3x2x3
  - 3 photo conditions
  - 2 grammar conditions
  - 3 price conditions
- 6 ad texts for 2 grammar conditions
  - But only 2 grammar conditions were the treatments of interest; we aggregate other variants by grammar condition

Reading Assignment

- Section 9.4 to page 307 in *Field Experiments*
- Reading notes
  - First two pages of this section about how to use regression to summarize the results of multifactor experiments
  - In Equation 9.10, the expression between the two equals signs just compact notation for the definition
    - $Y_i = Y_i(0)$, if d=0
    - $Y_i = Y_i(1)$, if d=1

Regression in Analysis of Multifactor Experiments

- Equation 9.11

$$Y_i = b_1 L_{\text{i Good Grammar}}^{\text{Non-Hispanic}} + b_2 L_{\text{i Good Grammar}}^{\text{Hispanic}} + b_3 L_{\text{i Bad Grammar}}^{\text{Non-Hispanic}} + b_4 L_{\text{i Bad Grammar}}^{\text{Hispanic}} + u_i$$

- Equation 9.12

$$Y_i = a + bJ_i + cG_i + d(J_i G_i) + u_i$$

- Very similar: fully saturated, with four parameters

Equation 9.11

$$Y_i = b_1 L_{\text{i Good Grammar}}^{\text{Non-Hispanic}} + b_2 L_{\text{i Good Grammar}}^{\text{Hispanic}} + b_3 L_{\text{i Bad Grammar}}^{\text{Non-Hispanic}} + b_4 L_{\text{i Bad Grammar}}^{\text{Hispanic}} + u_i$$

- Non-Hispanic, good grammar
- Hispanic, good grammar
- Non-Hispanic, bad grammar
- Hispanic, bad grammar

Equation 9.12

$$Y_i = a + bJ_i + cG_i + d(J_iG_i) + u_i$$

- Race of letter writer identified by J-indicator
  - J=1 for Jose
  - J=0 for Colin
- Grammar identified by G-indicator
  - G=1 for bad grammar
  - G=0 for good grammar
- Two equations equivalent though they look different
- Parameters related (see equations starting at end of p. 306)

Coefficients for Equation 9.11

$$Y_i = b_1 L_{\text{i Good Grammar}}^{\text{Non-Hispanic}} + b_2 L_{\text{i Good Grammar}}^{\text{Hispanic}} + b_3 L_{\text{i Bad Grammar}}^{\text{Non-Hispanic}} + b_4 L_{\text{i Bad Grammar}}^{\text{Hispanic}} + u_i$$

- Equation 9.11 measures the average within each of four cells.

	Colin	Jose
Good grammar	52%	37%
Bad grammar	29%	34%

- Coefficients b 1 , b 2 , b 3 , b 4 estimate the numbers from the table.

Estimating the Regression for Equation 9.11

- Define four dummy variables, one for each of the cells in the 2x2 table.

	Colin	Jose
Good grammar	52%	37%
Bad grammar	29%	34%

	Y	Non-Hispanic Good Grammar	Hispanic Good Grammar	Non-Hispanic Bad Grammar	Hispanic Bad Grammar
Recipient1	1				
Recipient2	0				
...					
Recipient400	1				

Coefficients for Equation 9.12

$$Y_i = a + bJ_i + cG_i + d(J_iG_i) + u_i$$

- In Equation 9.12, define two dummy variables, *J* and *G*.
- This specification measures differences between cells.
  - The intercept *a* measures the average response in the omitted category (Colin, no grammar errors).
  - The coefficient *b* tells us how much the ethnicity matters, for the case where there are no grammar errors (G=0).
  - The coefficient *c* tells us how much the grammar errors matter, for the case of Colin (J=0).
  - The coefficient *d* (interaction coefficient) tells us how much more the grammar errors matter for Jose than for Colin.

Interaction Coefficient in Equation 9.12

$$Y_i = a + bJ_i + cG_i + d(J_iG_i) + u_i$$

- Given when G=1 (the only time $d$matters).
  - What happens when J changes from 0 to 1?
- Regression can make the analysis simpler.
  - You can obtain results all at one time.
  - Results can be easier to interpret when coefficients are measuring differences instead of levels.

Estimated Coefficients for Equation 9.12

- Equation 9.16

$$\widehat{Y}_i = 0.52 + (-0.15)J_i + (-0.23)G_i + (0.20)(J_iG_i)$$

- These coefficients come from running a regression on the following data:

	Y	J	G	JG
Recipient1	1	1	1	1
Recipient2	0	1	1	1
...				
Recipient400	1	0	1	0

Original Data

	Colin	Jose
Good grammar	52%	37%
Bad grammar	29%	34%

- Remember: these cells show us exactly the coefficients obtained by running the regression in equation 9.11

Interpreting Estimated Coefficients for Equation 9.12

$$\widehat{Y}_i = 0.52 + (-0.15)J_i + (-0.23)G_i + (0.20)(J_i G_i)$$

- Regression coefficients found by subtracting numbers in the data table

	Colin	Jose
Good grammar	52%	37%
Bad grammar	29%	34%

- .52, the fraction of letters sent that received a response in that condition
- The second coefficient, -0.15
    - The difference between J=1 and J=0 when G=0
    - Additional response rate when changing letter writer from Colin to Jose
    - For the cases with good grammar (G=0)
    - .52 – .37 = – .15

Interaction Coefficient in Equation 9.12

$$\widehat{Y}_i = 0.52 + (-0.15)J_i + (-0.23)G_i + (0.20)(J_i G_i)$$

- How much more grammar treatment matters in Jose case vs. Colin case
- Difference in differences in the table

	Colin	Jose
Good grammar	52%	37%
Bad grammar	29%	34%

- .34 – .37 = – .03
- .29 – .52 = – .23
- .03 – (– .23) = .20

Interaction Coefficient in Equation 9.12 (cont.)

$$Y_i = a + bJ_i + cG_i + d(J_i G_i) + u_i$$

	Colin	Jose
Good grammar	52%	37%
Bad grammar	29%	34%

- How much more does bad grammar (G=1) matter with J=1 vs. J=0?
- How much more does the Jose treatment (J=1) matter with G=1 vs. G=0?
  - .34 – .29 = .05
    - This is how much more the J-treatment matters when G=1.
  - .37 – .52 = – .15
  - .5 – (– .15) = .20

Presenting Regression Results

- Regression output automatically includes standard errors, for easy hypothesis testing.
  - "Do grammar errors have less impact when letters are received from Jose vs. Colin?"
  - One can perform a t-test on the regression coefficient (.20 divided by the standard error).
  - The authors performed an F-test (unnecessarily complicated).
- Regression also gracefully handles non-binary categorical variables.
  - One example is the amount of someone's schooling.
  - Another example is a categorical variable that is more than binary.

What to Remember from This Week

- Bad controls: Don't include post-treatment covariates in regression
- Robust standard errors always a good idea
- Use clustered standard errors in regressions for experiments that have clustered treatment assignment
- Multifactor experiments: more than one treatment at a time
  - Regression with dummy variables quickly summarizes treatment effects and interaction effects

Points to Remember From This Week

- Quadratic terms in regression allow us to assess whether effects are increasing or decreasing in a continuous covariate.
- Attenuation bias: When X is mismeasured, our estimated treatment effect will be biased toward zero.
- Bad controls: In experiments, don't use covariates that could have been affected by the treatment.
- Robust standard errors are always a good idea. We don't have to know how the variance of the error term changes with covariates; the formula takes care of this for us.
- Clustered standard errors are necessary when we have clustered treatment assignment (see FE 3.6.2).
- Multifactor experiments have more than one treatment at a time.
    - We can use regression with dummy variables to quickly and conveniently summarize the treatment effects and interaction effects.