ECON 1123 Section 2

Slides at github.com/cjleggett/1123-section

Outline

- Name Circle
- Pset Feedback
- Lecture Recap / Questions
- Examples + Practice

Name Circle

Name Circle

- Name
- What’s one great class you’ve taken at Harvard?

Problem Set Feedback

What should I submit?

- If using Stata or just R:
  - File with all code (PDF, word doc, google doc…)
  - File with code (.R, .do)
  - File with code output (.log, .txt)
- If using R Markdown or Python Notebooks
  - .Rmd or .ipynb file with all code + answers
  - PDF export of all code + answers

What should I submit?

- If using Stata or just R:
  - **File with all code (PDF, word doc, google doc…)**
  - File with code (.R, .do)
  - File with code output (.log, .txt)
- If using R Markdown or Python Notebooks
  - .Rmd or .ipynb file with all code + answers
  - **PDF export of all code + answers**

General Submission Stuff

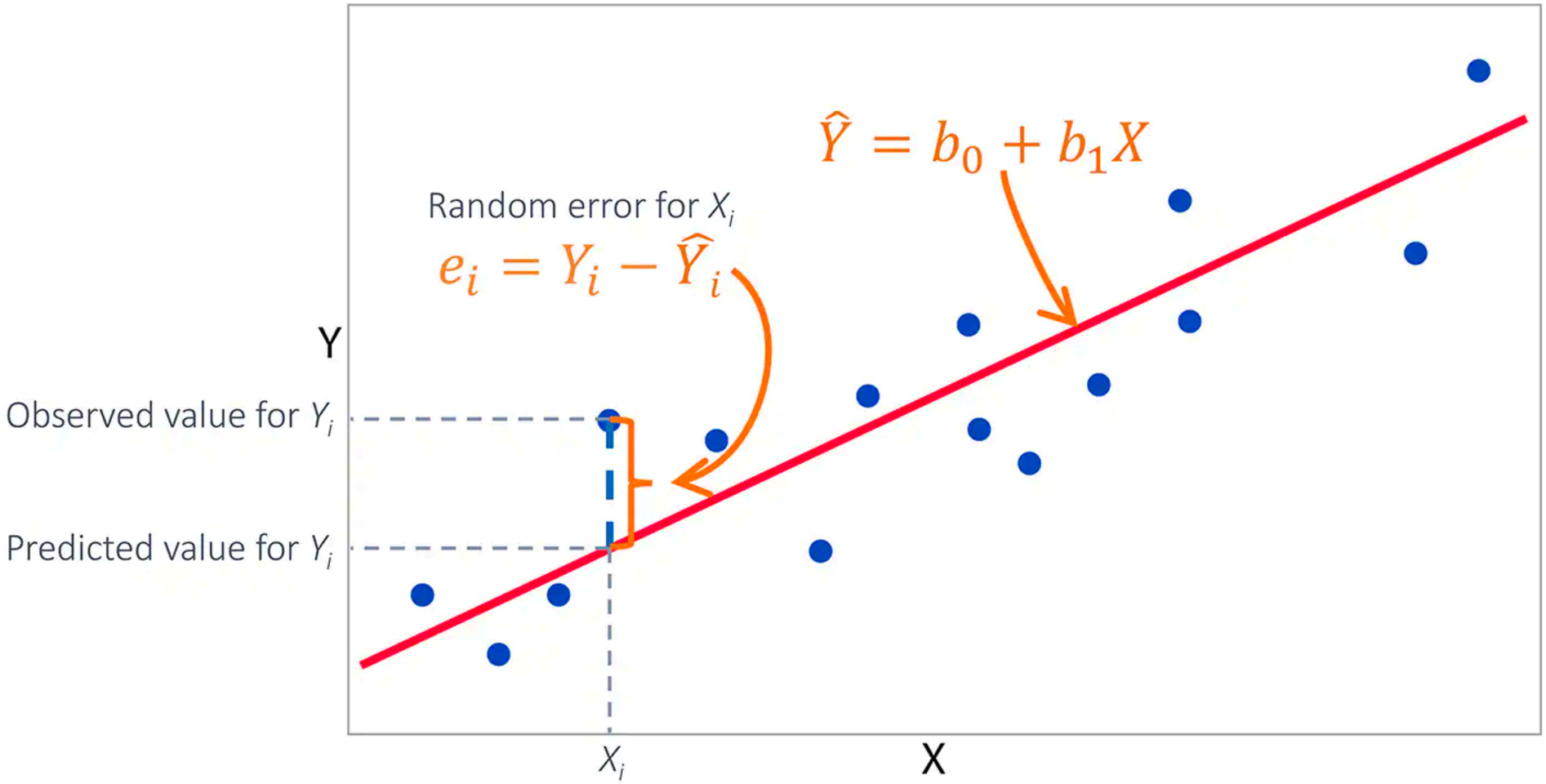
- Your PDF / word doc / google doc with your answers should have **everything** we need to grade (tables, images, calculations, graphs, etc). It’s fine if these are screenshots.
- Please assign your pages!!!
- Keep answers shorter!

Specific to this pset

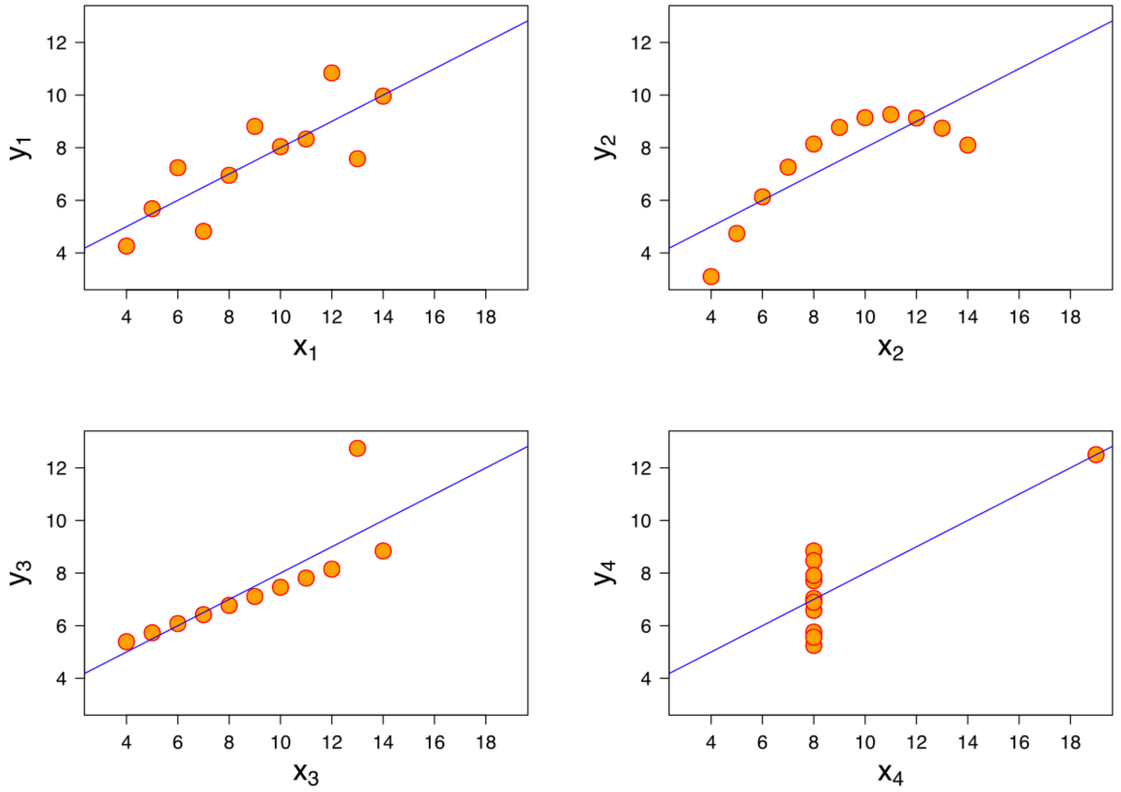
- **ALWAYS** use heteroskedasticity robust SEs in this class
  - Make sure you're also using them in tables and CIs
- In explanation questions, hit on all points mentioned!
- be very precise with definitions (discussion of 95% CI)

Lecture Recap

Least Squares



Anscombe: Data viz is important



Multiple Regression

- Pretty much most of what we do in this class!
- Allows us to control for other variables
- Same minimization problem, but using linear algebra to solve instead of algebra

How do we visualize this?

- Residuals!
  - Regress Y on x2, x3, x4… then make predictions, then use them to find residuals
  - Regress x1 on x2, x3, x4… then make predictions, then use them to find residuals
  - regress the residuals of Y on the residuals of x1!

Control for Indicator variables

- If we have N categories, how many indicator variables do we need to separate the groups?
- What does a0 represent?
- What's the difference in smoking between the south and midwest?

$$Smoke_i = \alpha_0 + \alpha_1 Midwest_i + \alpha_2 South_i + \alpha_3 West_i + \widetilde{Smoke_i}$$

Perfect Multicolinearity

- What about when we have too many variables?
- R + Stata will drop a variable and run a regression on the remaining ones
- Python will give a warning or throw an error

Heteroskedasticity

- Now we don’t have just two groups with different variances
- Do we still need to adjust for heteroskedasticity?
- YES! Always!
- Think about income vs distance from city center. Let’s try to graph it

Causal Effect

- Want to get at the effect of a treatment / condition
- This is impossible, but we can **estimate** the **average** effect
- Ideally a randomized control trial (RCT)
- These are often expensive, infeasible, or unethical :(
- So sometimes the best we can do is control for other variables

Omitted Variable Bias (OVB)

- Maybe the most important thing we'll learn in this class!
- No matter how many variables we have in our regression, we could always add more! (control for more things!)
- Economists reading and writing papers love to point out this missing (omitted) variables
- While we can't (always) know the true effect of the missing variables, we can guess its direction!

Fun Example: Let's make one together!

- Consider the "long" regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 U_i + u_i$$

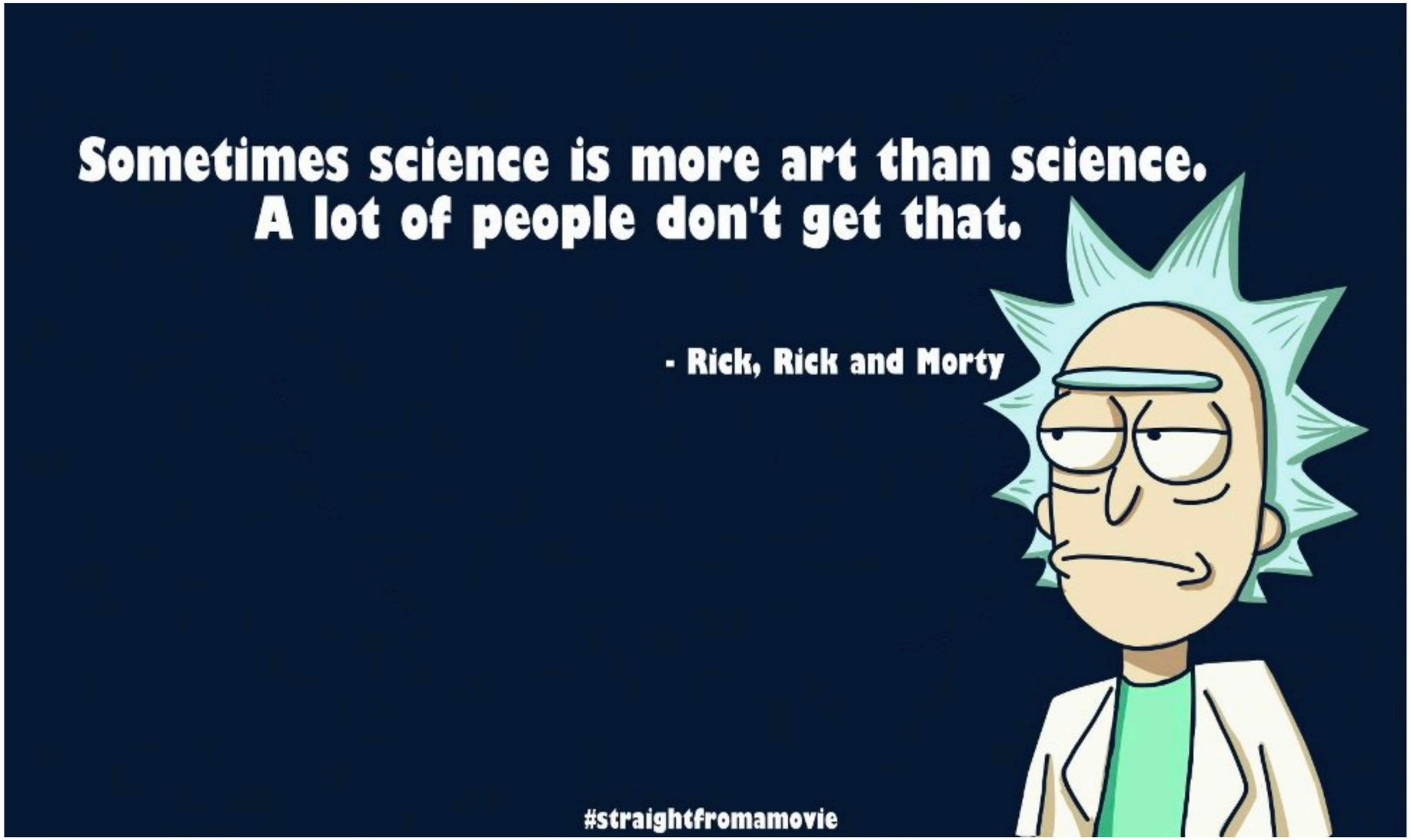
and the "short" regression:

$$Y_i = \alpha_0 + \alpha_1 X_i + v_i$$

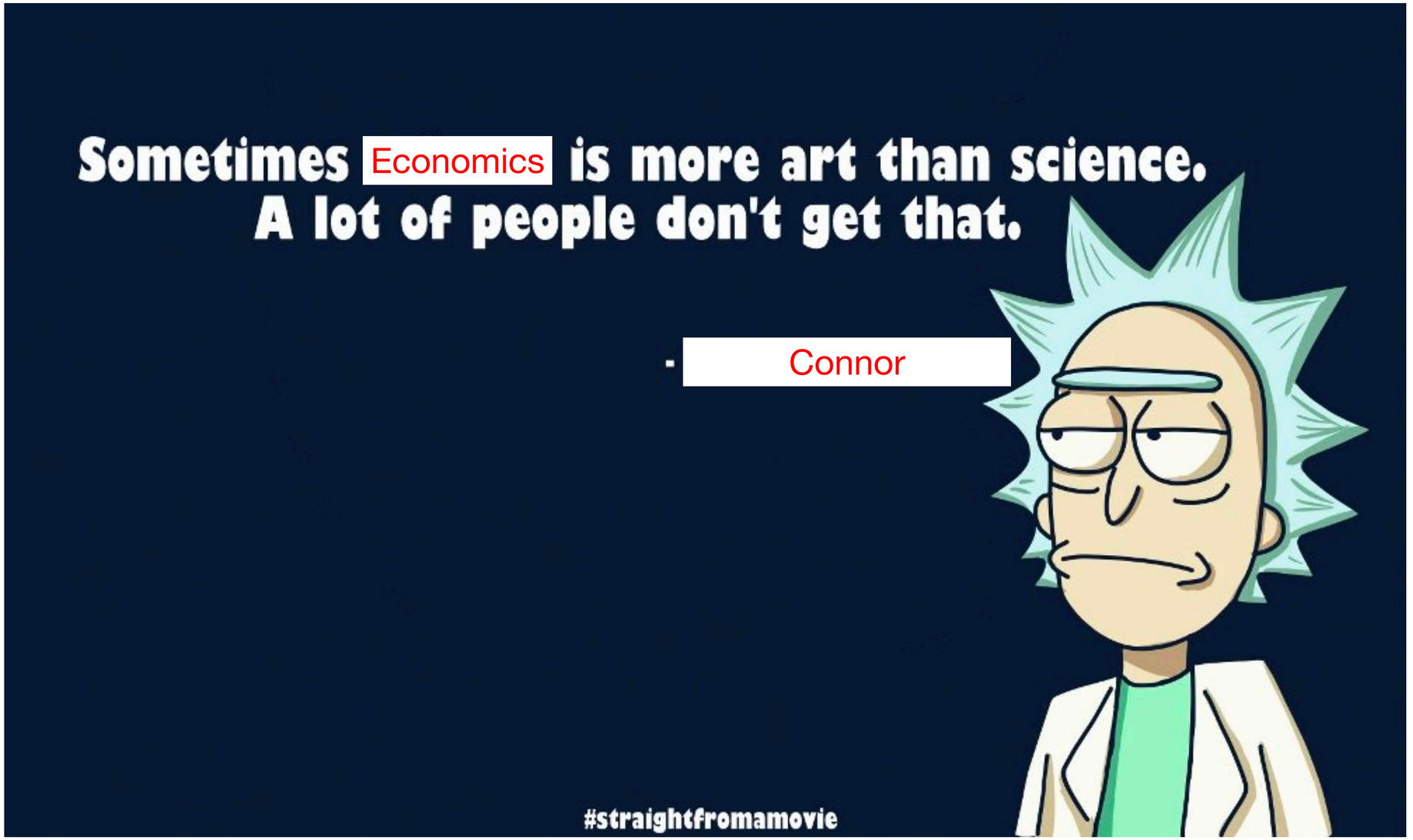
- What is the relationship between β_1 and α_1?
- Define the "auxiliary" regression:

$$U_i = \gamma_0 + \gamma_1 X_i + v_i$$

How can we really estimate Causality?



How can we really estimate Causality?



Exercises!