

Final review

Soc 500: Applied Social Statistics

Shay O'Brien, but mostly Simone Zhang and Ian Lundberg

Princeton University

January 15, 2019

Today's Agenda

- Learning from samples
- Bias-variance tradeoff with simulation and robust estimation
- Linear regression interpretation
- OLS assumptions, diagnostics, fixes
- Causal inference with observed variables
- Causal inference with unobserved confounders - Instrumental variables
- What if my code has a bug?
- Questions!

Sampling Distributions, Estimators, Point and Interval Estimation: Big ideas

- Sampling distributions
- Properties of estimators: evaluating estimators
- Point estimation and interval estimation (confidence intervals)

The problem

Suppose that you find yourself in front of this claw machine with an immense tank filled with PokeBalls



You want to know the average quality of the Pokemon in this machine. A sign tells you that the quality of the Pokemon in this machine is normally distributed with mean μ and variance σ^2 (but do not tell you what μ and σ^2 are).

Candidate estimators

You have a bit of pocket change to draw some PokeBalls from the machine. You want to decide how to make inferences about average quality based on the PokeBalls you draw, so you consider the following estimators for μ :

- $\hat{\mu}_1 = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
- $\hat{\mu}_2 = c$, where c is a constant
- $\hat{\mu}_3 = \frac{\sum_{i=1}^n X_i}{n+1}$

How do you decide which estimator to use?

Estimator Properties

Concept	Criteria	Intuition
Unbiasedness	$E[\hat{\mu}] = \mu$	Right on average
Efficiency	$V[\hat{\mu}_1] < V[\hat{\mu}_2]$	Low variance
Consistency	$\hat{\mu}_n \xrightarrow{P} \mu$	Converge to estimand as $n \rightarrow \infty$
Asymptotic Normality	$\hat{\mu}_n \overset{\text{approx.}}{\sim} N(\mu, \frac{\sigma^2}{n})$	Approximately normal in large n

- Finite sample properties (unbiasedness, efficiency): what happens across repeated sampling given a fixed n ?
- Asymptotic properties (consistency, asymptotic normality): what happens as n grows large?

Sampling distribution

Sampling distributions two ways:

- Analytical derivation: using information provided in the problem, properties of expectation and variance
- Simulation (will cover in more depth later): get lots of estimates by simulating lots of different draws

Let's derive the sampling distributions

To the board!

- $\hat{\mu}_1 = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$
- $\hat{\mu}_2 = c$, where c is a constant
- $\hat{\mu}_3 = \frac{\sum_{i=1}^n X_i}{n+1}$

What do we know?

$$X_i \sim N(\mu, \sigma^2)$$

$$\text{so } E(X_i) = \mu \text{ and } \text{Var}(X_i) = \sigma^2$$

(Some) useful properties of expectations:

- $E(X + Y) = E(X) + E(Y)$
- $E(a) = a$
- $E(aX) = aE(X)$
- $E(aX + bY) = aE(X) + bE(Y)$

(Some) useful properties of variances:

- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$
Note $\text{Cov}(X,Y) = 0$ if X and Y are independent
- $\text{Var}(a) = 0$
- $\text{Var}(X + a) = \text{Var}(X)$
- $\text{Var}(aX) = a^2\text{Var}(X)$
- $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)$

The sampling distributions

- $\hat{\mu}_1 = \frac{\sum_{i=1}^n X_i}{n} \sim N(\mu, \frac{\sigma^2}{n})$
- $\hat{\mu}_2 = c \sim E(\hat{\mu}_2) = c, \text{Var}(\hat{\mu}_2) = 0$
- $\hat{\mu}_3 = \frac{\sum_{i=1}^n X_i}{n+1} \sim N(\frac{n}{n+1}\mu, \frac{n}{(n+1)^2}\sigma^2)$

Let's evaluate by bias:

- $E(\hat{\mu}_1) - \mu = \mu - \mu = 0$
- $E(\hat{\mu}_2) - \mu = c - \mu$
- $E(\hat{\mu}_3) - \mu = \frac{n}{n+1}\mu - \mu = \frac{-1}{n+1}\mu$

$\hat{\mu}_1$ is the only unbiased estimator.

Efficiency and Consistency

Let's evaluate by efficiency:

- $Var(\hat{\mu}_1) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$
- $Var(\hat{\mu}_2) = 0$
- $Var(\hat{\mu}_3) = \frac{n\sigma^2}{(n+1)^2}$

$$Var(\hat{\mu}_2) < Var(\hat{\mu}_3) < Var(\hat{\mu}_1)$$

What about consistency?

Key: consider what happens as n gets large. Does the bias go away as $n \rightarrow \infty$? Does $Var(\hat{\mu})$ approach zero as $n \rightarrow \infty$?

Interval Estimation - Confidence Intervals

General form of a two-sided $(1 - \alpha)\%$ confidence interval:

$$\text{Point Estimate} \pm \text{Critical Value}_{\alpha/2} \cdot \text{Standard Error}$$

What do we need to find?

- Point estimate
- Sampling distribution of the estimator (how is it distributed? what's the standard deviation?)
 - Find critical value using `qnorm()`, `qt()`, etc. depending on the distribution you want. For example, `qnorm(0.975)` for a two-tailed 95% CI when you have a normal distribution.

Examples of two-sided CI's

Sample mean when we have a large sample:

$$\bar{X} \pm z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

Difference-in-means:

$$\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}$$

Interpretation

Suppose you get a 95% confidence interval for the mean quality of the Pokemon in the machine of (12 points, 21 points).

How would you explain what this confidence interval represents?

Interpretation

Suppose you get a 95% confidence interval for the mean quality of the Pokemon in the machine of (12 points, 21 points).

How would you explain what this confidence interval represents?

If we drew the same number of pokeballs from the same distribution and constructed confidence intervals in the same way many times, 95 percent of those confidence intervals would contain the true mean Poke-quality.

It does not mean that there's a .95 probability that the true parameter value lies between these two particular values! A given CI either contains the true value or not.

- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff
- 3 Linear regression interpretation
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables
- 7 What if my code has a bug?

- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff**
- 3 Linear regression interpretation
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables
- 7 What if my code has a bug?

Simulation: Big ideas

Steps in simulation:

- ① Set up 1 simulation
- ② Enclose that in a function
- ③ Replicate the function lots of times

Simulation example: Robust regression vs. OLS

- Robust regression reduces the influence of influential points
- But it may yield biased estimates if those outliers are warranted
- Given a population, can we simulate the distribution of these two estimators for β ?

Making the population

(There is nothing here to study. This is just how we made the dataset, so that if you want to play with it you can use this code and have the data on your computer)

```
set.seed(08544)
N <- 10000
x <- rexp(N, rate = .001) / 1000
error <- rnorm(N, mean = 0, sd = exp(x / 2))
y <- exp(x / 2) + error
pop <- data.frame(y,x)
```

Steps in simulation:

- ① **Set up 1 simulation**
- ② Enclose that in a function
- ③ Replicate the function lots of times

One simulation

```
set.seed(12345)
```

One simulation

```
set.seed(12345)
samp <- sample_n(pop, size = 20)
```

One simulation

```
set.seed(12345)
samp <- sample_n(pop, size = 20)
lm.fit <- lm(y ~ x, data = samp)
```


One simulation

```
set.seed(12345)
samp <- sample_n(pop, size = 20)
lm.fit <- lm(y ~ x, data = samp)
mm.fit <- rlm(y ~ x, method = "MM",
              maxit = 40, data = samp)
```

One simulation

```
set.seed(12345)
samp <- sample_n(pop, size = 20)
lm.fit <- lm(y ~ x, data = samp)
mm.fit <- rlm(y ~ x, method = "MM",
              maxit = 40, data = samp)
beta.lm <- coef(lm.fit)[2]
beta.mm <- coef(mm.fit)[2]
```

One simulation

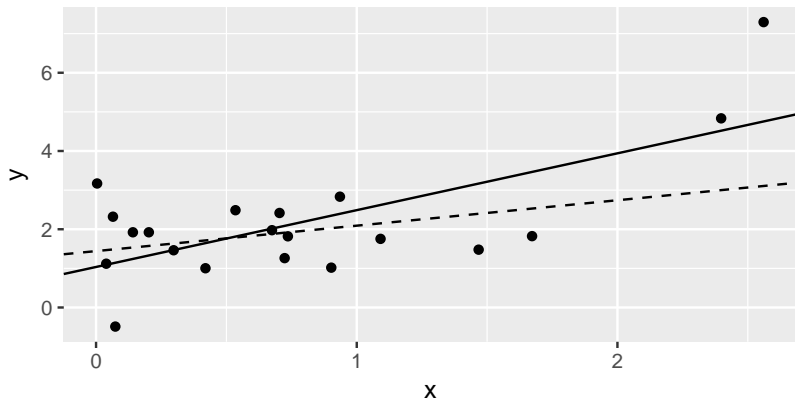
```
set.seed(12345)
samp <- sample_n(pop, size = 20)
lm.fit <- lm(y ~ x, data = samp)
mm.fit <- rlm(y ~ x, method = "MM",
             maxit = 40, data = samp)
beta.lm <- coef(lm.fit)[2]
beta.mm <- coef(mm.fit)[2]
c(beta.lm = beta.lm, beta.mm = beta.mm)

beta.lm.x beta.mm.x
1.4522595 0.6497666
```

One simulation

MM reduces the influence of the upper right points

OLS and MM fits
(MM is dashed)



Steps in simulation:

- ① Set up 1 simulation
- ② **Enclose that in a function**
- ③ Replicate the function lots of times

```
get.fit <- function(population) {  
  samp <- sample_n(pop, size = 20)  
  lm.fit <- lm(y ~ x, data = samp)  
  mm.fit <- rlm(y ~ x, method = "MM",  
               maxit = 40, data = samp)  
  beta.lm <- coef(lm.fit)[2]  
  beta.mm <- coef(mm.fit)[2]  
  return(c(beta.lm = beta.lm,  
          beta.mm = beta.mm))  
}
```

Steps in simulation:

- ① Set up 1 simulation
- ② Enclose that in a function
- ③ **Replicate the function lots of times**

```
set.seed(08544)
sims <- replicate(10000, get.fit(population = pop))
```


Properties of estimators, bias-variance tradeoff

Bias:

Properties of estimators, bias-variance tradeoff

Bias:

$$\text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$$

Properties of estimators, bias-variance tradeoff

Bias:

$$\text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$$

Variance:

$$\text{Var}(\hat{\beta})$$

Properties of estimators, bias-variance tradeoff

Bias:

$$\text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$$

Variance:

$$\text{Var}(\hat{\beta})$$

Mean squared error:

Properties of estimators, bias-variance tradeoff

Bias:

$$\text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$$

Variance:

$$\text{Var}(\hat{\beta})$$

Mean squared error:

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= E[(\hat{\beta} - \beta)^2] \\ &= \left(E[\hat{\beta} - \beta] \right)^2 + \text{Var}(\hat{\beta}) \\ &= \text{Bias}^2 + \text{Variance} \end{aligned}$$

Properties of estimators, bias-variance tradeoff

Bias:

$$\text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$$

Variance:

$$\text{Var}(\hat{\beta})$$

Mean squared error:

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= E[(\hat{\beta} - \beta)^2] \\ &= \left(E[\hat{\beta} - \beta] \right)^2 + \text{Var}(\hat{\beta}) \\ &= \text{Bias}^2 + \text{Variance} \end{aligned}$$

- Mean squared error is a measure of predictive validity.

Properties of estimators, bias-variance tradeoff

Bias:

$$\text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$$

Variance:

$$\text{Var}(\hat{\beta})$$

Mean squared error:

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= E[(\hat{\beta} - \beta)^2] \\ &= \left(E[\hat{\beta} - \beta] \right)^2 + \text{Var}(\hat{\beta}) \\ &= \text{Bias}^2 + \text{Variance} \end{aligned}$$

- Mean squared error is a measure of predictive validity.
- It is affected by both bias and variance.

Properties of estimators, bias-variance tradeoff

Bias:

$$\text{Bias}(\hat{\beta}) = E(\hat{\beta}) - \beta$$

Variance:

$$\text{Var}(\hat{\beta})$$

Mean squared error:

$$\begin{aligned} \text{MSE}(\hat{\beta}) &= E[(\hat{\beta} - \beta)^2] \\ &= \left(E[\hat{\beta} - \beta] \right)^2 + \text{Var}(\hat{\beta}) \\ &= \text{Bias}^2 + \text{Variance} \end{aligned}$$

- Mean squared error is a measure of predictive validity.
- It is affected by both bias and variance.
- In some cases, we are willing to choose a biased estimator because it has lower variance, and thus lower mean squared error. This is the **bias-variance tradeoff**.

Evaluate bias and variance

```
## Note the truth
```

```
truth <- coef(lm(y ~ x, data = pop))[2]
```

Evaluate bias and variance

```
## Note the truth
truth <- coef(lm(y ~ x, data = pop))[2]
## Look at the bias of the estimators
bias <- apply(sims, 1, mean) - truth
```

Evaluate bias and variance

```
## Note the truth
truth <- coef(lm(y ~ x, data = pop))[2]
## Look at the bias of the estimators
bias <- apply(sims, 1, mean) - truth
## Variance of the estimators
variance <- apply(sims, 1, var)
```

Evaluate bias and variance

```
## Note the truth
truth <- coef(lm(y ~ x, data = pop))[2]
## Look at the bias of the estimators
bias <- apply(sims, 1, mean) - truth
## Variance of the estimators
variance <- apply(sims, 1, var)
## Mean squared error
errors <- sims - truth
```

Evaluate bias and variance

```
## Note the truth
truth <- coef(lm(y ~ x, data = pop))[2]
## Look at the bias of the estimators
bias <- apply(sims, 1, mean) - truth
## Variance of the estimators
variance <- apply(sims, 1, var)
## Mean squared error
errors <- sims - truth
mse <- apply(errors, 1, function(x) mean(x ^ 2))
```


Evaluate bias and variance

```
## Note the truth
truth <- coef(lm(y ~ x, data = pop))[2]
## Look at the bias of the estimators
bias <- apply(sims, 1, mean) - truth
## Variance of the estimators
variance <- apply(sims, 1, var)
## Mean squared error
errors <- sims - truth
mse <- apply(errors, 1, function(x) mean(x ^ 2))
```

	Bias	Variance	Mean squared error
OLS	-0.19	2.88	2.92
MM estimation	-0.58	1.03	1.37

The MM estimator is biased but has lower variance and lower MSE.

```
## Plot the distribution of each estimator  
t(sims) %>%
```

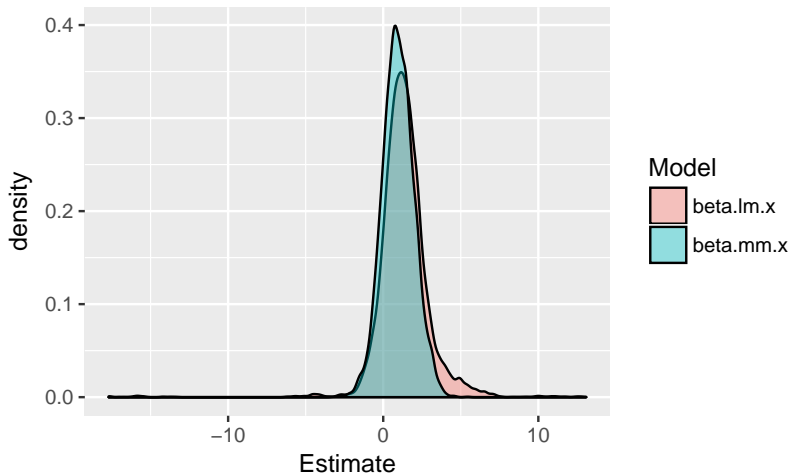


```
## Plot the distribution of each estimator
t(sims) %>%
  melt(value.name = "Estimate") %>%
```

```
## Plot the distribution of each estimator
t(sims) %>%
  melt(value.name = "Estimate") %>%
  mutate(Model = Var2) %>%
```

```
## Plot the distribution of each estimator
t(sims) %>%
  melt(value.name = "Estimate") %>%
  mutate(Model = Var2) %>%
  ggplot(aes(x = Estimate, fill = Model)) +
```

```
## Plot the distribution of each estimator
t(sims) %>%
  melt(value.name = "Estimate") %>%
  mutate(Model = Var2) %>%
  ggplot(aes(x = Estimate, fill = Model)) +
  geom_density(alpha = .4)
```



- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff
- 3 Linear regression interpretation
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables
- 7 What if my code has a bug?

- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff
- 3 Linear regression interpretation**
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables
- 7 What if my code has a bug?

Linear regression interpretation: Big ideas

- Interpreting coefficients
- Test difference in coefficients
- Interaction terms
- Plotting an interaction

Scenario

- You have a tutoring business.

Scenario

- You have a tutoring business.
- You want to show with data that your business is effective.

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$.

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT.

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT. **Incorrect.**

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT. **Incorrect.**
- ② Each day spent studying increases one's SAT score by 50 points.

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT. **Incorrect.**
- ② Each day spent studying increases one's SAT score by 50 points. **Incorrect.**

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT. **Incorrect.**
- ② Each day spent studying increases one's SAT score by 50 points. **Incorrect.**
- ③ Each day spent studying is associated with a 50 point increase in SAT score.

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT. **Incorrect.**
- ② Each day spent studying increases one's SAT score by 50 points. **Incorrect.**
- ③ Each day spent studying is associated with a 50 point increase in SAT score. **Better, but still not great.**
- ④ We expect that a student who spends 7 days studying will score 50 points higher than a student who spends 6 days studying.

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT. **Incorrect.**
- ② Each day spent studying increases one's SAT score by 50 points. **Incorrect.**
- ③ Each day spent studying is associated with a 50 point increase in SAT score. **Better, but still not great.**
- ④ We expect that a student who spends 7 days studying will score 50 points higher than a student who spends 6 days studying. **Incorrect.**
- ⑤ We expect that a student who spends 7 days studying will score 50 points higher than a student who spends 6 days studying, assuming they either both have a tutor or both

Interpreting coefficients

Suppose we fit a model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 300$. Which is/are a correct non-causal interpretation of $\hat{\beta}_1$?

- ① Every additional day spent studying causes one to score 50 points higher on the SAT. **Incorrect.**
- ② Each day spent studying increases one's SAT score by 50 points. **Incorrect.**
- ③ Each day spent studying is associated with a 50 point increase in SAT score. **Better, but still not great.**
- ④ We expect that a student who spends 7 days studying will score 50 points higher than a student who spends 6 days studying. **Incorrect.**
- ⑤ We expect that a student who spends 7 days studying will score 50 points higher than a student who spends 6 days studying, assuming they either both have a tutor or both

Test difference in coefficients

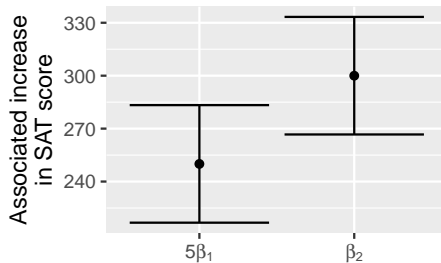
$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

- A critic of your business says, “Tutoring is no more beneficial than spending 5 days studying.”

Test difference in coefficients

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

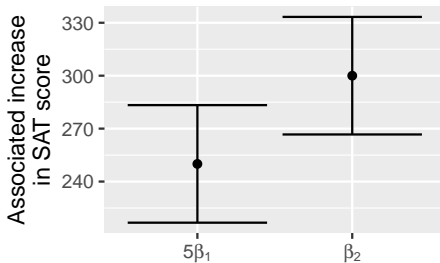
- A critic of your business says, “Tutoring is no more beneficial than spending 5 days studying.”
- You produce this graph:



Test difference in coefficients

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) + u$$

- A critic of your business says, “Tutoring is no more beneficial than spending 5 days studying.”
- You produce this graph:



- Have you disproven your critic?

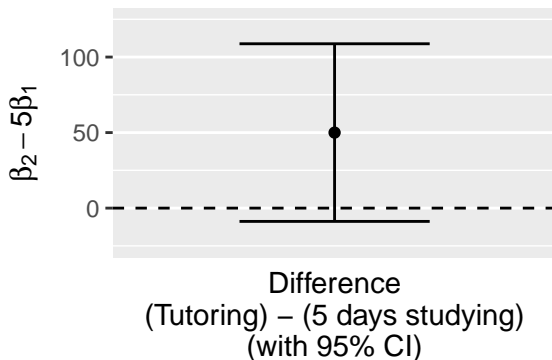
No. You need explicitly look at uncertainty about the difference.

Test difference in coefficients

- A critic of your business says, “Tutoring is no more beneficial than spending 5 days studying.”

Test difference in coefficients

- A critic of your business says, “Tutoring is no more beneficial than spending 5 days studying.”
- Now you produce this graph:



- Have you disproven your critic?

No. 0 is within the 95% confidence interval, so we cannot reject the null hypothesis that tutoring and 5 days studying are equally beneficial. The test is inconclusive.

You think your critic is totally wrong, since studying is more beneficial if you have a tutor!

You think your critic is totally wrong, since studying is more beneficial if you have a tutor!

Your paid consultant proposes the model:

You think your critic is totally wrong, since studying is more beneficial if you have a tutor!

Your paid consultant proposes the model:

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Days studying})(\text{Having a tutor}) + u$$

Do you like this model?

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) \\ & + \beta_2(\text{Days studying})(\text{Having a tutor}) + u \end{aligned}$$

We don't like this model. It assumes that

$$\begin{aligned} E(\text{SAT} \mid 0 \text{ days studying, tutor}) &= E(\text{SAT} \mid 0 \text{ days studying, no tutor}) \\ &= \beta_0 \end{aligned}$$

This is an unreasonable assumption! When we fit a model with an interaction, it should always include the lower-order terms. We need a term for having a tutor.

Interaction terms

Suppose we fit a new model

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ & + \beta_3(\text{Having a tutor})(\text{Days studying}) + u \end{aligned}$$

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **without a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **without a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.
 - 50 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **without a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.
 - 50 point higher SAT score.
 - 75 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **without a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.
 - 50 point higher SAT score.
 - 75 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **without a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.
 - 50 point higher SAT score.
 - 75 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **with a tutor**, we expect that a student who studies one more day than another student will have a:

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **with a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **with a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.
 - 50 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **with a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.
 - 50 point higher SAT score.
 - 75 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- Among students **with a tutor**, we expect that a student who studies one more day than another student will have a:
 - 25 point higher SAT score.
 - 50 point higher SAT score.
 - 75 point higher SAT score.

Interaction terms

Suppose we fit a new model

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ & + \beta_3(\text{Having a tutor})(\text{Days studying}) + u \end{aligned}$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- How do we interpret $\hat{\beta}_3 = 25$?

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- How do we interpret $\hat{\beta}_3 = 25$?
 - 1 Among those with tutors, each day studying increases SAT score by 25 points.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- How do we interpret $\hat{\beta}_3 = 25$?
 - ① Among those with tutors, each day studying increases SAT score by 25 points.
 - ② Among those with tutors, we expect that students who study one more day than other students will have SAT scores that are 25 points higher.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- How do we interpret $\hat{\beta}_3 = 25$?
 - ① Among those with tutors, each day studying increases SAT score by 25 points.
 - ② Among those with tutors, we expect that students who study one more day than other students will have SAT scores that are 25 points higher.
 - ③ The positive association between having a tutor and SAT score is 25 points greater for each additional day of studying.

Interaction terms

Suppose we fit a new model

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

We estimate $\hat{\beta}_1 = 50$, $\hat{\beta}_2 = 100$, $\hat{\beta}_3 = 25$. Which is/are a correct non-causal interpretation of these results?

- How do we interpret $\hat{\beta}_3 = 25$?
 - ① Among those with tutors, each day studying increases SAT score by 75 points.
 - ② Among those with tutors, we expect that students who study one more day than other students will have SAT scores that are 75 points higher.
 - ③ The positive association between having a tutor and SAT score is 25 points greater for each additional day of studying.

Plotting interactions

Your model:

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ & + \beta_3(\text{Having a tutor})(\text{Days studying}) + u \end{aligned}$$

Plotting interactions

Your model:

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ & + \beta_3(\text{Having a tutor})(\text{Days studying}) + u \end{aligned}$$

How could we convey this interaction in a plot?

Plotting interactions

Your model:

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

How could we convey this interaction in a plot?

① (go to board)

Plotting interactions

Your model:

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

How could we convey this interaction in a plot?

- 1 (go to board)
- 2 What would the y -axis be?

Plotting interactions

Your model:

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

How could we convey this interaction in a plot?

- 1 (go to board)
- 2 What would the y -axis be?
- 3 The x -axis?

Plotting interactions

Your model:

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

How could we convey this interaction in a plot?

- 1 (go to board)
- 2 What would the y -axis be?
- 3 The x -axis?
- 4 How does having a tutor appear on the plot?

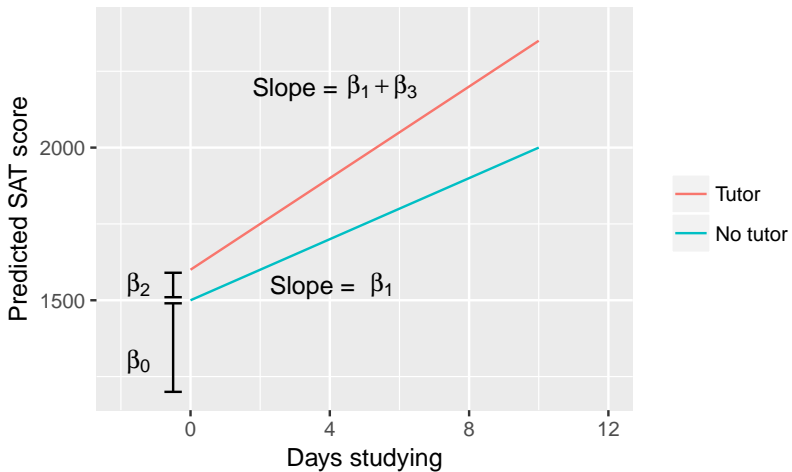
Plotting interactions

Your model:

$$\text{SAT score} = \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ + \beta_3(\text{Having a tutor})(\text{Days studying}) + u$$

How could we convey this interaction in a plot?

- ① (go to board)
- ② What would the y -axis be?
- ③ The x -axis?
- ④ How does having a tutor appear on the plot?
- ⑤ Where is β_0 ? β_1 ? β_2 ? β_3 ?



Plotting interactions with lots of variables

Your model:

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ & + \beta_3(\text{Having a tutor})(\text{Days studying}) \\ & + \beta_4(\text{Air temperature on SAT day}) \\ & + \beta_5(\text{Number of sharpened pencils}) + u \end{aligned}$$

Plotting interactions with lots of variables

Your model:

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ & + \beta_3(\text{Having a tutor})(\text{Days studying}) \\ & + \beta_4(\text{Air temperature on SAT day}) \\ & + \beta_5(\text{Number of sharpened pencils}) + u \end{aligned}$$

- How could we convey this interaction in a plot?

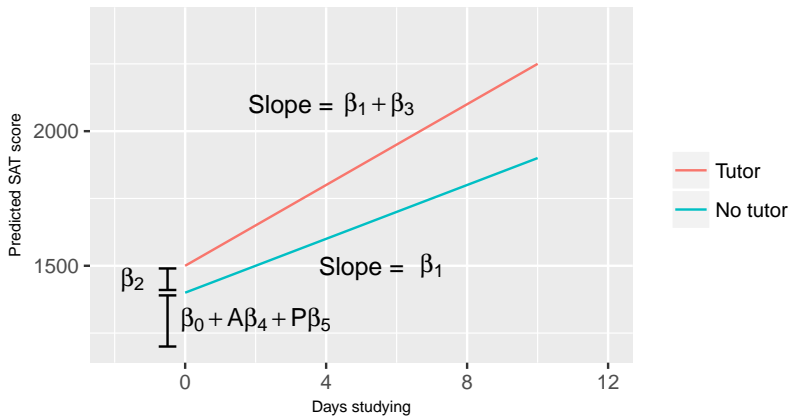
Plotting interactions with lots of variables

Your model:

$$\begin{aligned} \text{SAT score} = & \beta_0 + \beta_1(\text{Days studying}) + \beta_2(\text{Having a tutor}) \\ & + \beta_3(\text{Having a tutor})(\text{Days studying}) \\ & + \beta_4(\text{Air temperature on SAT day}) \\ & + \beta_5(\text{Number of sharpened pencils}) + u \end{aligned}$$

- How could we convey this interaction in a plot?
- We could set the other covariates to their means when making predicted values.

Predicted values holding air temperature and sharpened pencils at their means
(A = mean air temperature, P = mean # sharpened pencils)



Linear regression interpretation: Big ideas

- Interpreting coefficients
- Test difference in coefficients
- Interaction terms
- Plotting an interaction

OLS Assumptions, Diagnostics, and Fixes: Big ideas

- OLS assumptions: what they mean, why we need them
- Strategies for diagnosing OLS assumption violations and correcting for them

General OLS Assumptions

- ① Linearity: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$
- ② Random/iid sample: (y_i, \mathbf{x}'_i) are a iid sample from the population.
- ③ No perfect collinearity: \mathbf{X} is an $n \times (K + 1)$ matrix with rank $K + 1$
- ④ Zero conditional mean: $\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{0}$
- ⑤ Homoskedasticity: $\text{var}(\mathbf{u}|\mathbf{X}) = \sigma_u^2 \mathbf{I}_n$
- ⑥ Normality: $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_n)$

Assumptions Needed for Unbiasedness and Consistency

Linearity The population regression model is linear in its parameters and correctly specified

- Result: biased/inconsistent estimates
- Diagnose: scatterplots with LOESS lines, added variable plot, component residual plots, GAMs (is the edf for the smooth term > 1)?
- Correct: transformations, polynomials, different models

Assumptions Needed for Unbiasedness and Consistency

Random/iid sample The observed data represent a random sample from the population described by the model. This can be violated if we have correlation across observations (spatial, multiple observations from same cluster, same unit at multiple times)

- Result: no bias in parameters with appropriate alternative assumptions (structured dependence)
- Result (ii): violations imply heteroskedasticity
- Result (iii): outliers from different distributions can cause inefficiency/bias
- Diagnose: plotting residuals, checking out extreme values, Durbin-Watson for serial correlation
- Correct: including fixed effects/random effects, aggregate data to cluster level, cluster-robust standard errors, HAC standard errors for serial correlation, etc.

Extreme values

Extreme values could be a violation of iid (those observations could come from a different distribution altogether).

Reminder about the three types:

- Outlier: extreme in Y
- Leverage point: extreme in X
- Influence point: extreme in X and Y

Two strategies:

- Identifying and removing if appropriate
- Robust estimation (eg MM estimation)

Identifying extreme values

- **Outlier:** Studentized residuals. In R: `rstudent(model)`
- **Leverage:** Hat values. In R: `hatvalues(model)`
- **Influence point:**
 - Standardized influence of a given observation on a particular coefficient: `DFbetaS`. In R: `dfbetas(model)`
 - Influence across all coefficients: Cook's distance

Reminder: the built-in diagnostics can be helpful: `plot(model)`

Assumptions Needed for Unbiasedness and Consistency

No perfect collinearity No explanatory variable is the linear combination of the other explanatory variables

- Result: We don't get an estimate!
- Diagnose/correct: drop one collinear term
- Note we also cannot get an estimate if we have no variation in X .

Zero conditional mean Expected value of the error term is zero conditional on all values of the explanatory variable: $E(u | X) = 0$. Note it implies that $\text{Cov}(X, u) = 0$.

- Result: biased/inconsistent estimates
- Diagnose: very difficult. Ask yourself: are there any unobserved confounders?
- Correct: instrumental variables and other methods that deal unmeasured confounding Note: Not to be confused with the mean of the residuals!

Homoskedasticity

Needed for large-sample inference: The error term has the same variance conditional on all values of the explanatory variable.

- Result: Our estimates are unbiased but our standard errors are biased
- Diagnostics:
 - Visual: plot residuals vs fitted values (is there fanning?), spread-location plots
 - Formal tests: Breusch-Pagan, Cook-Weisberg, etc.
- Correcting: transform dependent variable, adjust using weights, **use an estimator of $\text{Var}(\hat{\beta})$ that's robust to heteroskedasticity**, admit we have the wrong model

Heteroskedasticity Consistent Estimator

- Under non-constant error variance:

$$\text{Var}[\mathbf{u}] = \mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

- When $\mathbf{\Sigma} \neq \sigma^2 \mathbf{I}$, we are stuck with this expression:

$$\text{Var}[\hat{\beta}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Sigma}\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

- Idea: If we can consistently estimate the components of $\mathbf{\Sigma}$, we could directly use this expression by replacing $\mathbf{\Sigma}$ with its estimate, $\hat{\mathbf{\Sigma}}$.

White's Heteroskedasticity Consistent Estimator

Suppose we have **heteroskedasticity of unknown form**:

$$V[\mathbf{u}] = \mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

then $V[\hat{\beta}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{\Sigma}\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$ and White (1980) shows that

$$\widehat{V[\hat{\beta}|\mathbf{X}]} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \begin{bmatrix} \hat{u}_1^2 & 0 & 0 & \dots & 0 \\ 0 & \hat{u}_2^2 & 0 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & \hat{u}_n^2 \end{bmatrix} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

is a consistent estimator of $V[\hat{\beta}|\mathbf{X}]$ **under any form of heteroskedasticity** consistent with $V[\mathbf{u}]$ above.

The estimate based on the above is called the **heteroskedasticity consistent (HC)** or **robust standard errors**.

White's Heteroskedasticity Consistent Estimator

Robust standard errors are easily computed with the “sandwich” formula:

- ① Fit the regression and obtain the residuals $\hat{\mathbf{u}}$
- ② Construct the “meat” matrix $\hat{\Sigma}$ with squared residuals in diagonal:

$$\hat{\Sigma} = \begin{bmatrix} \hat{u}_1^2 & 0 & 0 & \dots & 0 \\ 0 & \hat{u}_2^2 & 0 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & \hat{u}_n^2 \end{bmatrix}$$

- ③ Plug $\hat{\Sigma}$ into the sandwich formula to obtain the robust estimator of the variance-covariance matrix

$$V[\hat{\beta}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\hat{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

- There are various **small sample corrections** to improve performance when sample size is small. The most common variant (sometimes labeled HC1) is:

$$V[\hat{\beta}|\mathbf{X}] = \frac{n}{n-k-1} \cdot (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\hat{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Regular & Robust Standard Errors in R

```
> library(sandwich)
> library(lmtest)
> # Homoskedasticity:
> coeftest(mod1)
> # Classic White
> coeftest(mod1,vcov = vcovHC(mod1, type = "HC0"))
> # Small sample correction
> coeftest(mod1,vcov = vcovHC(mod1, type = "HC1"))
```

Note: You can pass coeftest into stargazer to print out nice regression results with robust standard errors.

Normality of errors

Needed for inference in small samples: The error term is independent of the explanatory variables and normally distributed.

- Result: critical values for t and F tests wrong
- Diagnose: checking the (studentized) residuals, QQ-plots
If the errors are Normal, the studentized residuals follow a t distribution with $(n - k - 2)$ degrees of freedom
- Correct: transformations, add variables to X, different model

Instrumental Variables - The Big Ideas

- Identifying assumptions - what are they? Do they hold up?
- Estimation - Wald, 2SLS
- Interpreting results - what have we identified?

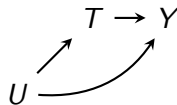
Causal effect of juvenile detention

Let's work through this example:

Aizer, Anna, and Joseph Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics* 759-804.

Research Question

- Question: What is the causal effect of juvenile detention on high school completion and adult incarceration?
- What are possible unobserved confounders? In what direction might those confounders bias our results?



The Natural Experiment

The initial judge that a child encounters is randomly assigned. It is “a function of the sequence with which cases happen to enter into the system and the judge availability that is set in advance.” 96% of people who come before the court are found guilty, but judges exercise discretion over whether a child is placed on probation or detained. This paper exploits variation in judge sentencing behaviour.

- Data: 10 years of administrative data from the Chicago Public Schools Student Database, Juvenile Court of Cook County Delinquency Database, Illinois Department of Corrections Adult Admissions and Exits Database

What's the identification strategy?

- **What's the instrument (Z)?**

What's the identification strategy?

- **What's the instrument (Z)?**

Being assigned a judge with a particular propensity to incarcerate (calculated based on all other cases before the judge)

What's the identification strategy?

- **What's the instrument (Z)?**
Being assigned a judge with a particular propensity to incarcerate (calculated based on all other cases before the judge)
- **What's the treatment (T)?**

What's the identification strategy?

- **What's the instrument (Z)?**
Being assigned a judge with a particular propensity to incarcerate (calculated based on all other cases before the judge)
- **What's the treatment (T)?**
Being incarcerated as a juvenile

What's the identification strategy?

- **What's the instrument (Z)?**
Being assigned a judge with a particular propensity to incarcerate (calculated based on all other cases before the judge)
- **What's the treatment (T)?**
Being incarcerated as a juvenile
- **What are the outcomes (Y)?**

What's the identification strategy?

- **What's the instrument (Z)?**
Being assigned a judge with a particular propensity to incarcerate (calculated based on all other cases before the judge)
- **What's the treatment (T)?**
Being incarcerated as a juvenile
- **What are the outcomes (Y)?**
1) High school completion, 2) adult incarceration

Let's evaluate the assumptions

1 Exogeneity of the Instrument

Let's evaluate the assumptions

① Exogeneity of the Instrument

Are judges randomly assigned?

Let's evaluate the assumptions

- ① **Exogeneity of the Instrument**
Are judges randomly assigned?
- ② **Exclusion Restriction**

Let's evaluate the assumptions

① Exogeneity of the Instrument

Are judges randomly assigned?

② Exclusion Restriction

Can a judge's propensity to incarcerate affect high school completion/adult incarceration through causal channels other than through juvenile incarceration?

Let's evaluate the assumptions

① Exogeneity of the Instrument

Are judges randomly assigned?

② Exclusion Restriction

Can a judge's propensity to incarcerate affect high school completion/adult incarceration through causal channels other than through juvenile incarceration?

③ First-stage relationship

Let's evaluate the assumptions

① Exogeneity of the Instrument

Are judges randomly assigned?

② Exclusion Restriction

Can a judge's propensity to incarcerate affect high school completion/adult incarceration through causal channels other than through juvenile incarceration?

③ First-stage relationship

Does a judge's propensity to incarcerate induce variation in sentences? (checkable by regressing treatment on the instrument)

④ Monotonicity

Let's evaluate the assumptions

① Exogeneity of the Instrument

Are judges randomly assigned?

② Exclusion Restriction

Can a judge's propensity to incarcerate affect high school completion/adult incarceration through causal channels other than through juvenile incarceration?

③ First-stage relationship

Does a judge's propensity to incarcerate induce variation in sentences? (checkable by regressing treatment on the instrument)

④ Monotonicity

Are there defiers?

Framework

Name	$T_i(1)$	$T_i(0)$	
Always Takers	1	1	Incarcerated regardless of judge assigned
Never Takers	0	0	Not incarcerated regardless of judge assigned
Compliers	1	0	Incarcerated if strict judge, but not if lenient judge
Defiers	0	1	Incarcerated if lenient judge, not if strict judge

- The potential outcomes for members of which groups differ based on judge assignment?

Results

”...juvenile detention is estimated to decrease high school graduation by 13 percentage points and increase adult incarceration by 23 percentage points”

Wait, but what effect did we identify?

We identified the **complier average treatment effect**.

What does this mean substantively? To whom does this effect generalize?

Wait, but what effect did we identify?

We identified the **complier average treatment effect**.

What does this mean substantively? To whom does this effect generalize?

We have estimated the effect of incarceration for marginal juvenile cases in which the sentencing outcome is affected by the judge assigned. We might believe that these effects might be different for children who commit particularly mild or particularly severe crimes.

- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff
- 3 Linear regression interpretation
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables
- 7 What if my code has a bug?

- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff
- 3 Linear regression interpretation
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables**
- 7 What if my code has a bug?

Causal inference with observed variables: Big ideas

- Define a causal quantity of interest
- Consider the experimental ideal
- Identification
- Estimation

Running example

Does marriage cause men to earn more per hour?

Defining the causal quantity of interest

What is the treatment? What is the outcome?

$$\tau_i = Y_i(1) - Y_i(0)$$

Defining the causal quantity of interest

What is the treatment? What is the outcome?

$$\tau_i = Y_i(1) - Y_i(0)$$

It is impossible to observe both potential outcomes $Y_i(1)$ and $Y_i(0)$.

Defining the causal quantity of interest

What is the treatment? What is the outcome?

$$\tau_i = Y_i(1) - Y_i(0)$$

It is impossible to observe both potential outcomes $Y_i(1)$ and $Y_i(0)$.

Example:

The causal effect of marriage on a man's hourly wage is the difference between his potential wage if he were married vs. if he were not.

The experimental ideal

What experiment would answer this question?

The experimental ideal

What experiment would answer this question?

Example:

If we randomly assigned some men to marry and others to not, we could compare the two groups.

Identification

If we had an infinite number of observations, would we recover the ATE?

(see Morgan and Winship, p. 78-79)

Identification

If we had an infinite number of observations, would we recover the ATE?

(see Morgan and Winship, p. 78-79)

Example:

- If marriage was randomly assigned:

Identification

If we had an infinite number of observations, would we recover the ATE?

(see Morgan and Winship, p. 78-79)

Example:

- If marriage was randomly assigned: identified.

Identification

If we had an infinite number of observations, would we recover the ATE?

(see Morgan and Winship, p. 78-79)

Example:

- If marriage was randomly assigned: identified.
- If we just compared married and unmarried men:

Identification

If we had an infinite number of observations, would we recover the ATE?

(see Morgan and Winship, p. 78-79)

Example:

- If marriage was randomly assigned: identified.
- If we just compared married and unmarried men: not identified.

Identification

If we had an infinite number of observations, would we recover the ATE?

(see Morgan and Winship, p. 78-79)

Example:

- If marriage was randomly assigned: identified.
- If we just compared married and unmarried men: not identified.
- More assumptions needed

Assumption: Selection on observables

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i \mid X_i$$

Example:

Net of education and age, marriage is independent of potential wages.

Assumption: Selection on observables

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i \mid X_i$$

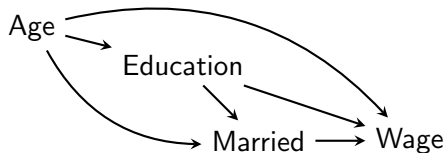
Example:

Net of education and age, marriage is independent of potential wages.

Can we draw in a DAG?

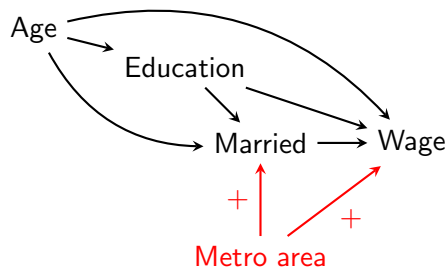
DAGs: Confounders

Net of education and age, marriage is independent of potential wages.



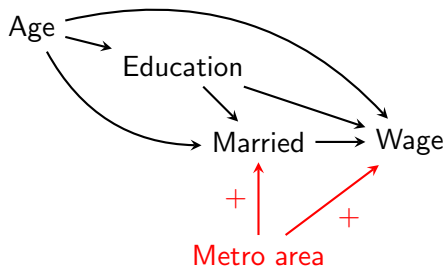
Sensitivity and expected direction of bias

Potential violation: Which way is the estimated effect of T on Y biased?



Sensitivity and expected direction of bias

Potential violation: Which way is the estimated effect of T on Y biased?

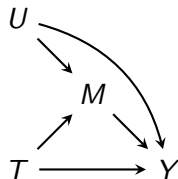


Answer: Upwardly biased (+ times + = +).

For every critic, you can try to reason about the bias.

DAGs: Colliders

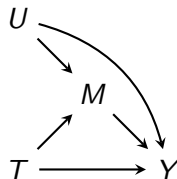
Theory (example next slide):



- Which of these variables is a collider? On what path?

DAGs: Colliders

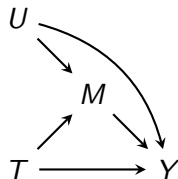
Theory (example next slide):



- Which of these variables is a collider? On what path?
- M is a collider on the path $T \rightarrow M \leftarrow U \rightarrow Y$

DAGs: Colliders

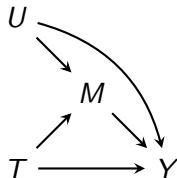
Theory (example next slide):



- Which of these variables is a collider? On what path?
- M is a collider on the path $T \rightarrow M \leftarrow U \rightarrow Y$
- What happens if we ignore M ?

DAGs: Colliders

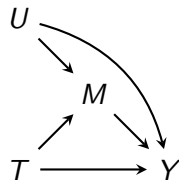
Theory (example next slide):



- Which of these variables is a collider? On what path?
- M is a collider on the path $T \rightarrow M \leftarrow U \rightarrow Y$
- What happens if we ignore M ? (Nothing: $T \rightarrow Y$ is identified!)

DAGs: Colliders

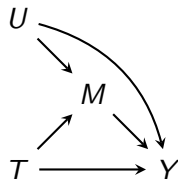
Theory (example next slide):



- Which of these variables is a collider? On what path?
- M is a collider on the path $T \rightarrow M \leftarrow U \rightarrow Y$
- What happens if we ignore M ? (Nothing: $T \rightarrow Y$ is identified!)
- What happens if we condition on M ?

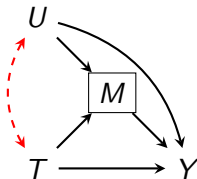
DAGs: Colliders

Theory (example next slide):



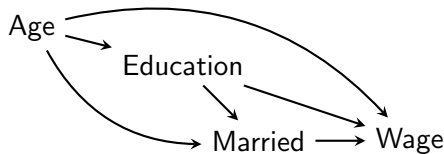
- Which of these variables is a collider? On what path?
- M is a collider on the path $T \rightarrow M \leftarrow U \rightarrow Y$
- What happens if we ignore M ? (Nothing: $T \rightarrow Y$ is identified!)
- What happens if we condition on M ? Bad things: Induced association!)

DAGs: Induced association from conditioning on a collider



Back to our example

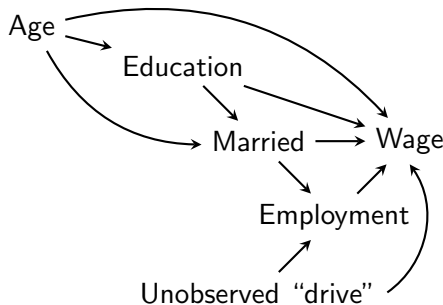
Net of education and age, marriage is independent of potential wages.



DAGs: Collider example

Net of education and age, marriage is independent of potential wages.

But, our sample is only among those who are employed, since others don't have wages.

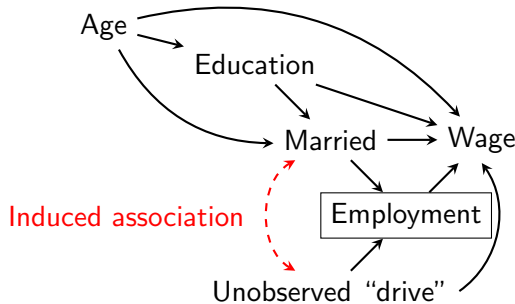


DAGs: Collider example

Net of education and age, marriage is independent of potential wages.

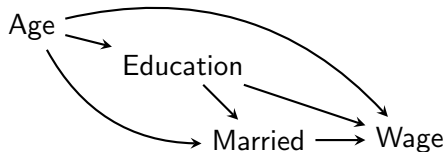
But, our sample is only among those who are employed, since others don't have wages.

This effectively conditions on employment, causing problems!



Estimation

Given the assumptions (selection on observables) in the DAG below, we have identified the effect of marriage on men's wages.



To **estimate** this effect, we will assume an additive linear regression model.

$$Y = \beta_0 + \beta_1(\text{Age}) + \beta_2(\text{Education}) + \beta_3(\text{Married}) + u$$

Causal inference with observed variables: Big ideas

- Define a causal quantity of interest
- Consider the experimental ideal
- Identification
- Estimation

- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff
- 3 Linear regression interpretation
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables
- 7 What if my code has a bug?

- 1 Learning From Samples
- 2 Simulation and the bias-variance tradeoff
- 3 Linear regression interpretation
- 4 OLS Assumptions, Diagnostics, Fixes
- 5 Instrumental Variables
- 6 Causal inference with observed variables
- 7 What if my code has a bug?

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?
- 3 Check for % signs that need to be \% in Latex math mode. These comment out the remainder of the line, which can make problems.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?
- 3 Check for % signs that need to be $\%$ in Latex math mode. These comment out the remainder of the line, which can make problems.
- 4 Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?
- 3 Check for % signs that need to be \% in Latex math mode. These comment out the remainder of the line, which can make problems.
- 4 Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.
- 5 Open a new RMarkdown file.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?
- 3 Check for % signs that need to be \% in Latex math mode. These comment out the remainder of the line, which can make problems.
- 4 Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.
- 5 Open a new RMarkdown file.
 - Paste the first part of your RMarkdown in it.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?
- 3 Check for % signs that need to be \% in Latex math mode. These comment out the remainder of the line, which can make problems.
- 4 Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.
- 5 Open a new RMarkdown file.
 - Paste the first part of your RMarkdown in it.
 - See if you can knit.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?
- 3 Check for % signs that need to be \% in Latex math mode. These comment out the remainder of the line, which can make problems.
- 4 Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.
- 5 Open a new RMarkdown file.
 - Paste the first part of your RMarkdown in it.
 - See if you can knit.
 - Add some more.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- 1 Stay calm.
- 2 Look at what you just typed. Does anything seem odd?
- 3 Check for % signs that need to be $\%$ in Latex math mode. These comment out the remainder of the line, which can make problems.
- 4 Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.
- 5 Open a new RMarkdown file.
 - Paste the first part of your RMarkdown in it.
 - See if you can knit.
 - Add some more.
 - See if you can knit.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- ① Stay calm.
- ② Look at what you just typed. Does anything seem odd?
- ③ Check for % signs that need to be \% in Latex math mode. These comment out the remainder of the line, which can make problems.
- ④ Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.
- ⑤ Open a new RMarkdown file.
 - Paste the first part of your RMarkdown in it.
 - See if you can knit.
 - Add some more.
 - See if you can knit.
 - Continue until you find the bug.

Dealing with bugs in code:

Knit your PDF constantly. Then you'll know when you type something that has a bug.

If your file won't knit:

- ① Stay calm.
- ② Look at what you just typed. Does anything seem odd?
- ③ Check for % signs that need to be \% in Latex math mode. These comment out the remainder of the line, which can make problems.
- ④ Check for unusual characters (like β) that you might have copied and pasted from the exam PDF.
- ⑤ Open a new RMarkdown file.
 - Paste the first part of your RMarkdown in it.
 - See if you can knit.
 - Add some more.
 - See if you can knit.
 - Continue until you find the bug.

For R errors, also check that you didn't forget to load an R package (they need to be loaded within the RMarkdown file, not by hand in the console).