PSC 202
SYRACUSE UNIVERSITY
INTRODUCTION
TO POLITICAL
ANALYSIS
MULTIPLE REGRESSION, PART 1

## REMINDERS

- Problem Set 8 due on Friday
- Section worksheet also due on Friday
- Graded pass/fail, counts towards section attendance/participation
- If you have questions about the material, please email and/or attend student hours


## WHERE WE ARE

- Formulate research question
- Propose explanation/theory, hypotheses
- Data collection process
- Use data to evaluate hypotheses
- Reassess explanation


WHERE WE ARE

Sleep (X)


## GPA (Y)

- Fundamental problem of causal inference
- Causal effect of sleep: GPA if you sleep a lot - GPA if you sleep little
- We don't observe you in both states, only in one (either you sleep a lot, or not)


## ZERO-ORDER RELATIONSHIP

## How much sleep during finals?

More Than 6<br>Hours/Night

6 Or Fewer<br>Hours/Night

Average Gpa


- Frequency in parentheses

Sleep (X)


## GPA (Y)

- Fundamental problem of causal inference
- We can't compare a student's GPA when they sleep more vs. when they sleep less (only one is observed)
- We can compare GPA of students who sleep a lot with GPA of students who sleep little
- But: Students who sleep a lot might be different from students who don't in many other ways
- And these other differences might affect GPA as well

Number of Classes (Z)

More classes $=$ less sleep



More classes $=$ worse grades

GPA (Y)

Could be spurious

- Does sleep still have effect on GPA when controlling for number of classes?


## CONTROLLED EFFECTS

5 Or Fewer Classes

6 Or More Classes

Sleep
More Than 6 Or Fewer More Than 6 Or Fewer Hours/Night Hours/Night Hours/Night Hours/Night

Average Gpa

| 3.49 |
| :--- |
| $(20)-0.02(34)$ |


| 3.78 | 3.63 |
| :--- | ---: |
| $(20)$ | $0.15(18)$ |

- Idea: Look at relation between sleep and GPA among people who take many classes and (separately) among people who take few classes


## CONTROLLED EFFECTS

## 5 Or Fewer Classes

6 Or More Classes

Sleep
More Than 66 Or Fewer More Than 6 Or Fewer Hours/Night Hours/Night Hours/Night Hours/Night

Average
Gpa


- Still differences in GPA between more/less sleep when controlling for number of classes.


## CONTROLLED EFFECTS

5 Or Fewer Classes

6 Or More Classes

Sleep
More Than 6 Or Fewer More Than 6 Or Fewer Hours/Night Hours/Night Hours/Night Hours/Night

Average Gpa

| 3.78 | 3.63 |
| :--- | ---: |
| $(20)$ | $0.15(18)$ |

- More nuance: Interactive relation. Sleep matters a lot for GPA among people who take many classes, sleep matters less for GPA among people who take fewer classes


# Anxiety (Z) 

More anxious = less sleep


More anxious = worse grades

GPA (Y)

Could be spurious

- Does sleep still have effect on GPA when controlling for how anxious students are?


## CONTROLLED EFFECTS

Anxiety: Below 7

Anxiety: 7 Or Above

Average Gpa


- Idea: Look at relation between sleep and GPA among people who are less anxious and (separately) among people who are more anxious


## CONTROLLED EFFECTS

Anxiety: Below 7

Anxiety: 7 Or Above

$$
\begin{array}{l|r}
\hline 3.75 & 3.50 \\
\hline(18) 0.25(26)
\end{array}
$$

$$
\begin{array}{l|r}
\hline 3.54 & 3.6 \\
\hline(22)-0.06(27)
\end{array}
$$

- Sleep still matters for GPA after controlling for how anxious people are


## HURDLES TO CAUSALITY

- Is there a credible causal mechanism that connects X to Y ?
- Can we rule out the possibility that $Y$ could cause X ?
- Is there covariation between $X$ and $Y$ ?
- Have we controlled for allconfounding variables $(Z)$ that might make the association between $X$ and $Y$ spurious?


## EXAMPLE



## 2 INDEPENDENT VARIABLES

More Anxious
Less Anxious

| Sleep More | Sleep Less | Sleep More | Sleep Less |
| :--- | :--- | :--- | :--- |
| Gpa |  |  |  |

## 3 INDEPENDENT VARIABLES

## More Classes

## Fewer Classes

More Anxious Less Anxious More Anxious Less Anxious

| Sleep More | Sleep Less | Sleep More | Sleep Less | Sleep More | Sleep Less | Sleep More | Sleep Less |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Gpa |  |  |  |  |  |  |  |

## 4 INDEPENDENT VARIABLES

## Male

## Female

More Classes Fewer Classes More Classes Fewer Classes

|  | More <br> Anxious |  | Less <br> Anxious |  | More <br> Anxious |  | Less <br> Anxious |  | More <br> Anxious |  | Less <br> Anxious |  | More Anxious |  | Less <br> Anxious |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Sleep <br> More | Sleep Less | Sleep More | Sleep Less | Sleep More | Sleep Less | Sleep <br> More | $\begin{array}{\|c} \text { Sleep } \\ \text { Less } \end{array}$ | Sleep <br> More | Sleep Less | Sleep <br> More | Sleep Less | Sleep <br> More | Sleep Less | Sleep <br> More | Sleep |
| Gpa |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

- As we control for more potential confounders, table gets increasingly unwieldy (and few/no observations in some cells)


## ANOTHER ISSUE

- What if a control is interval-level?
- e.g. sleep or number of classes
- I divided into more/less sleep and more/fewer classes, but that's not ideal
- But if I didn't do this, unwieldy table with many cells where there are no observations
- Multiple regression
- Extends bivariate regression to incorporate not just one, but many independent variables
- Logic of controlled comparison:
- Separate students with more classes and fewer classes
- Among students w/ more classes: What is the effect of sleep on GPA?
- Among students w/ fewer classes: What is the effect of sleep on GPA?
- Gives us the partial effects of sleep, holding number of classes constant


## MULTIPLE REGRESSION

- Multiple regression does something similar
- Can estimate the effect of two variables on dependent variable (GPA)
- Gives the partial effect of sleep, holding number of classes constant
- And: gives the partial effect of number of classes, holding sleep constant


## MULTIPLE REGRESSION

- Can include more than 2 independent variables
- e.g. sleep, number of classes, how anxious
- Gives the partial effect of sleep, holding constant classes and anxiety
- Gives the partial effect of classes, holding constant sleep and anxiety
- Gives the partial effect of anxiety, holding constant sleep and classes


## BIVARIATE REGRESSION



## BIVARIATE REGRESSION

Coefficient Standard Error T-Value

Sleep: Hours/
Night

| 3.35 | 0.16 | 20.80 |
| :--- | :--- | :--- |
| 0.04 | 0.03 | 1.52 |

## LINEAR REGRESSION

- $y=a+b_{1}{ }^{*} x_{1}$
- y: GPA
- $\mathrm{x}_{1}$ : hours of sleep/night


## LINEAR REGRESSION

- Let's add number of classes as a second control
- $y=a+b_{1}{ }^{*} x_{1}+b_{2}{ }^{*} x_{2}$
- y: GPA
- $x_{1}$ : hours of sleep/night
- $x_{2}$ : number of classes


## EFFECT OF SLEEP

## Coefficient Standard Error T-Value

| Intercept | 3.60 | 0.28 | 8.74 |
| :---: | :---: | :---: | :---: |
| Sleep: Hours/ <br> Night | 0.04 | 0.03 | 1.46 |
| Number Of <br> Classes | 0.14 | 0.05 | 2.96 |

EFFECT OF SLEEP

- Coefficient: 0.04 (SE 0.03, t-value 1.46)
- Interpretation: For every additional hour of sleep, GPA increases by 0.04 points, holding all other variables constant
- We cannot reject $H_{0}$, so effect of sleep on GPA is not statistically significant at the $5 \%$ level


## EFFECT OF NUMBER OF CLASSES

Coefficient

Standard Error
T-Value

Intercept

Sleep: Hours/ Night

Number Of
Classes
$0.14 \quad 0.05$
2.96

EFFECT OF AGE

- Coefficient: 0.14 (SE 0.05, t-value 2.96)
- Interpretation: For every additional class taken, GPA increases by 0.14 points, holding all other variables constant
- We can reject $\mathrm{H}_{0}$, so effect of number of classes on GPA is significant at the $5 \%$ level


## INTERCEPT

## Coefficient Standard Error T-Value



## INTERCEPT

- Intercept: 3.60
- Gives expected GPA when both sleep=0 and number of classes=0


## LINEAR REGRESSION

- Let's add anxiety as another control
- $y=a+b_{1}{ }^{*} x_{1}+b_{2}{ }^{*} x_{2}+b_{3}{ }^{*} x_{3}$
- y: GPA
- $x_{1}$ : hours of sleep/night
- $\mathrm{x}_{2}$ : number of classes
- $x_{3}$ : anxiety score (1-10, higher values=more anxious)


## INTERCEPT

## Coefficient Standard Error T-Value

| Intercept | 2.69 | 0.31 | 8.59 |
| :---: | :---: | :---: | :---: |
| Sleep: Hours/ <br> Night | 0.04 | 0.03 | 1.49 |
| Number Of <br> Classes | 0.14 | 0.05 | 2.91 |
| Anxiety | -0.01 | 0.01 | -0.96 |

CETERIS PARIBUS

- Linear regression allows us to estimate the effect of several independent variables on the dependent variable
- Gives us the effect of an independent variable on the dependent variable, holding all other variables constant
- "ceteris paribus"
- We can assess the effect of the variables independently of each other


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14^{*}$ Classes 0.01*Anxiety
- What is the predicted GPA for a person who sleeps 6 hours, takes 5 classes, and with an anxiety score of 7 ?


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14^{*}$ Classes 0.01*Anxiety
- What is the predicted GPA for a person who sleeps 6 hours, takes 5 classes, and with an anxiety score of 7 ?
- GPA $=2.96+0.04 * 6+0.14 * 5-0.01 * 7=3.83$


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14 *$ Classes 0.01*Anxiety
- What is that person slept 8 hours instead?


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14^{*}$ Classes 0.01*Anxiety
- What is that person slept 8 hours instead?
- GPA $=2.96+0.04 * 8+0.14 * 5-0.01 * 7=3.91$


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14 *$ Classes 0.01*Anxiety
- What is that person took 4 classes instead?


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14^{*}$ Classes 0.01*Anxiety
- What is that person took 4 classes instead?
- GPA $=2.96+0.04 * 8+0.14 * 4-0.01 * 7=3.77$


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14 *$ Classes 0.01*Anxiety
- And what if they had an anxiety score of 10 ?


## CETERIS PARIBUS

- GPA $=2.96+0.04 *$ Sleep $+0.14 *$ Classes 0.01*Anxiety
- And what if they had an anxiety score of 10 ?
- GPA $=2.96+0.04 * 8+0.14 * 4-0.01 * 10=3.74$


## CETERIS PARIBUS

- Of course, being more anxious might also lead to sleeping less
- But the linear regression estimates the effects of sleep and anxiety independently of each other


## LINEAR REGRESSION

- So far: The independent variables were interval-level
- Hours of sleep, number of classes, anxiety score (1-10)
- What if independent variable is nominal or ordinal-level?
- e.g. effect of gender


## DUMMY VARIABLE REGRESSION

- Nominal or ordinal-level independent variable can easily be incorporated in linear regression
- $y=a+b_{1}{ }^{*} x_{1}+b_{2}{ }^{*} x_{2}+b_{3}{ }^{*} x_{3}+b_{4}{ }^{*} x_{4}$
- $x_{3}=0$ if gender=female
- $x_{3}=1$ if gender=male
- Same idea as before, but $x_{4}$ can only take values of 0 or 1
- "Dummy variable"
- 0/1


## REGRESSION

|  | Coefficient | Standard Error | T-Value |
| :---: | :---: | :---: | :---: |
| Intercept | 2.75 | 0.33 | 8.38 |
| Sleep: Hours/ <br> Night | 0.04 | 0.03 | 1.50 |
| Number Of <br> Classes | 0.14 | 0.05 | 2.75 |
| Anxiety | -0.02 | 0.01 | -1.07 |
| Gender (Male) | -0.05 | 0.08 | -0.62 |

## EFFECT OF GENDER

- Coefficient: -0.05 (SE 0.08, t-value 0.62)
- Where female is coded 0 and male coded 1
- Interpretation: If someone is male, their GPA is expected to be 0.05 points lower than if someone is female, holding all other variables constant
- However, we do not reject $\mathrm{H}_{0}$, so effect of gender on GPA is not significant at the $5 \%$ level


## EFFECT OF GENDER

- GPA $=2.75+0.04 *$ Sleep $+0.14 *$ Classes 0.02*Anxiety $-0.05^{*}$ Gender
- female: gender=0
- male: gender=1


## EFFECT OF GENDER

- GPA $=2.75+0.04 *$ Sleep $+0.14^{*}$ Classes 0.02*Anxiety $-0.05^{*}$ Gender
- female: gender=0
- male: gender=1
- What is the predicted GPA for a student who sleeps 6 hours, takes 5 classes, has an anxiety score of 9 , and who is male?


## EFFECT OF GENDER

- GPA $=2.75+0.04 *$ Sleep $+0.14 *$ Classes 0.02*Anxiety $-0.05^{*}$ Gender
- female: gender=0
- male: gender=1
- What is the predicted GPA for a student who sleeps 6 hours, takes 5 classes, has an anxiety score of 9 , and who is male?
- GPA $=2.75+0.04 * 6+0.14 * 5-0.02 * 9-$
$0.05^{*} 1=3.46$


## EFFECT OF GENDER

- GPA $=2.75+0.04 *$ Sleep $+0.14^{*}$ Classes 0.02*Anxiety $-0.05^{*}$ Gender
- female: gender=0
- male: gender=1
- What is the predicted GPA for a student who sleeps 6 hours, takes 5 classes, has an anxiety score of 9 , and who is female?


## EFFECT OF GENDER

- GPA $=2.75+0.04 *$ Sleep $+0.14 *$ Classes 0.02*Anxiety $-0.05^{*}$ Gender
- female: gender=0
- male: gender=1
- What is the predicted GPA for a student who sleeps 6 hours, takes 5 classes, has an anxiety score of 9 , and who is female?
- GPA $=2.75+0.04 * 6+0.14 * 5-0.02 * 9-$
$0.05^{*} 0=3.51$


## SLIDERS AND SWITCHES

Categorical variables


## WHAT THIS ALLOWS US TO DO

- Multiple regression is a tool that allows us to tackle the fourth hurdle to causality
- Have we controlled for all confounding variables (Z) that might make the association between $X$ and $Y$ spurious?
- We can now estimate effect of $X$ on $Y$ controlling for all confounders we can think of ( $Z_{1}, Z_{2}$, etc.)


## WHAT THIS ALLOWS US TO DO

- If we have not one theory about what influences $Y$, but many theories, we can test which one's have an effect on $Y$ and which don't


## HOW IS THIS USEFUL?

- What causes high infant mortality rates?

- Infant mortality rates (Death under 1 year of age per 1,000 live births)

