GV300 - Quantitative Political Analysis

University of Essex - Department of Government

Lorenzo Crippa

Week 21 – 17 February, 2020

Today's session

Today we will mostly work in two directions:

Today's session

Today we will mostly work in two directions:

1. Some final remarks on the power of a test

Today's session

Today we will mostly work in two directions:

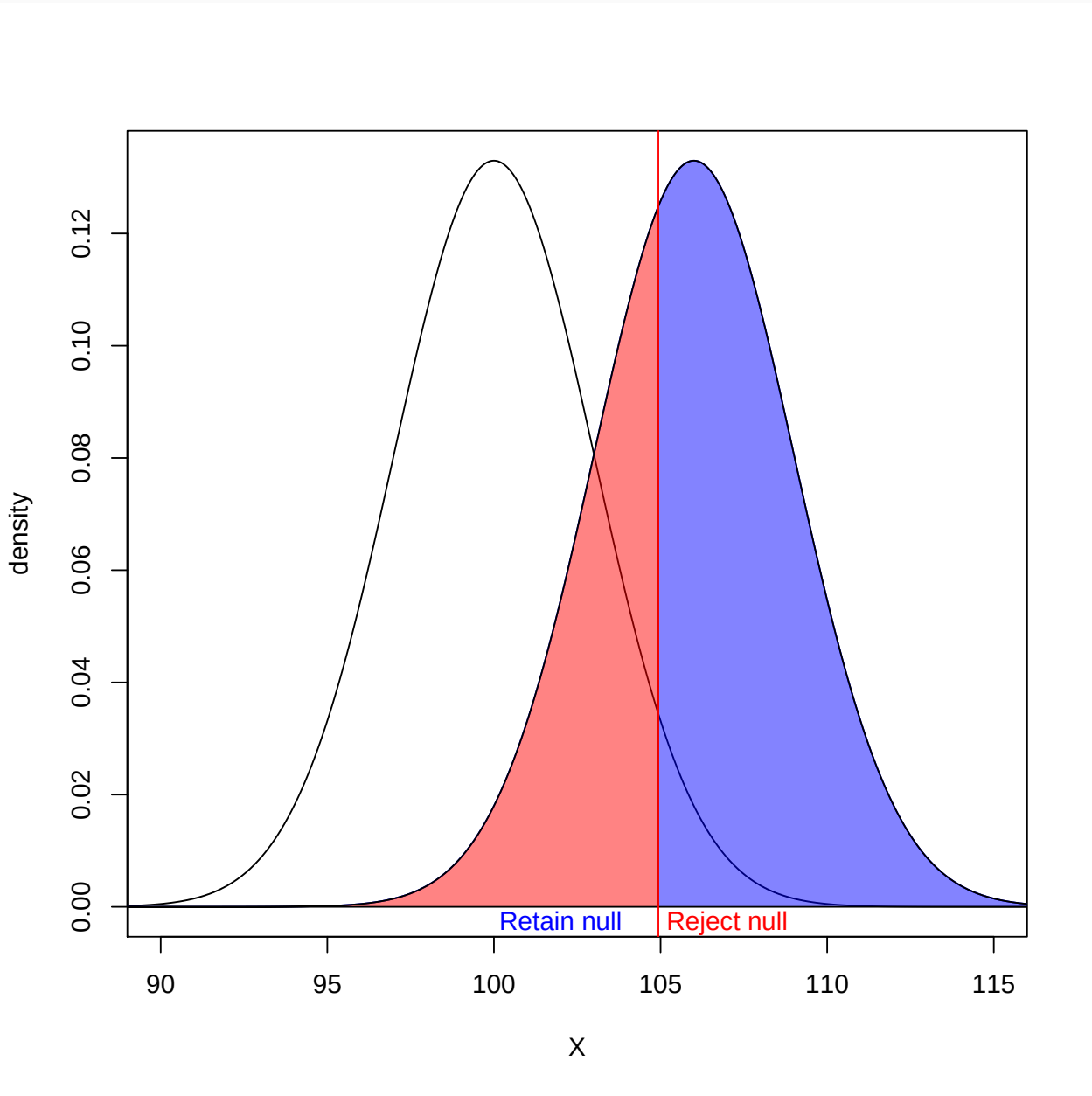
1. Some final remarks on the power of a test
2. An exercise on an observational dataset from Card (1993). Try to make a causal claim

Final remarks on the power of a test

Again on the power of a test

- LABa01: power analysis (see R file)
- LABa02: discussion on the power of a test, α, β and $1 - \beta$ (see R file)

The power of a test



Exercise on observational dataset from Card (1993)

R and Stata session

Today we are working on the replication data of an article by Card (1993). You can work alone or in groups of 2.

R and Stata session

Today we are working on the replication data of an article by Card (1993). You can work alone or in groups of 2.

1. Download the dataset `Card_data.dta` from Moodle.

R and Stata session

Today we are working on the replication data of an article by Card (1993). You can work alone or in groups of 2.

1. Download the dataset `Card_data.dta` from Moodle.
2. Open the dataset, look at the variables you have and try to understand what they measure

R and Stata session

Today we are working on the replication data of an article by Card (1993). You can work alone or in groups of 2.

1. Download the dataset `Card_data.dta` from Moodle.
2. Open the dataset, look at the variables you have and try to understand what they measure
3. Think about what research question you could answer by using these data. Find a causal claim you can test with them

Card's (1993) research question

We have a dataset on 3010 subjects, identified by an `id` variable.

Card's (1993) research question

We have a dataset on 3010 subjects, identified by an `id` variable. We have information on their socio-economic conditions, their education and demographics.

Card's (1993) research question

We have a dataset on 3010 subjects, identified by an `id` variable. We have information on their socio-economic conditions, their education and demographics.

The research question from the author is:

Card's (1993) research question

We have a dataset on 3010 subjects, identified by an `id` variable. We have information on their socio-economic conditions, their education and demographics.

The research question from the author is:

Do years of schooling affect wage?

Causality and problems

Think about the causal claim you hypothesize.

Causality and problems

Think about the causal claim you hypothesize.

1. Draw a causal diagram representing it

Causality and problems

Think about the causal claim you hypothesize.

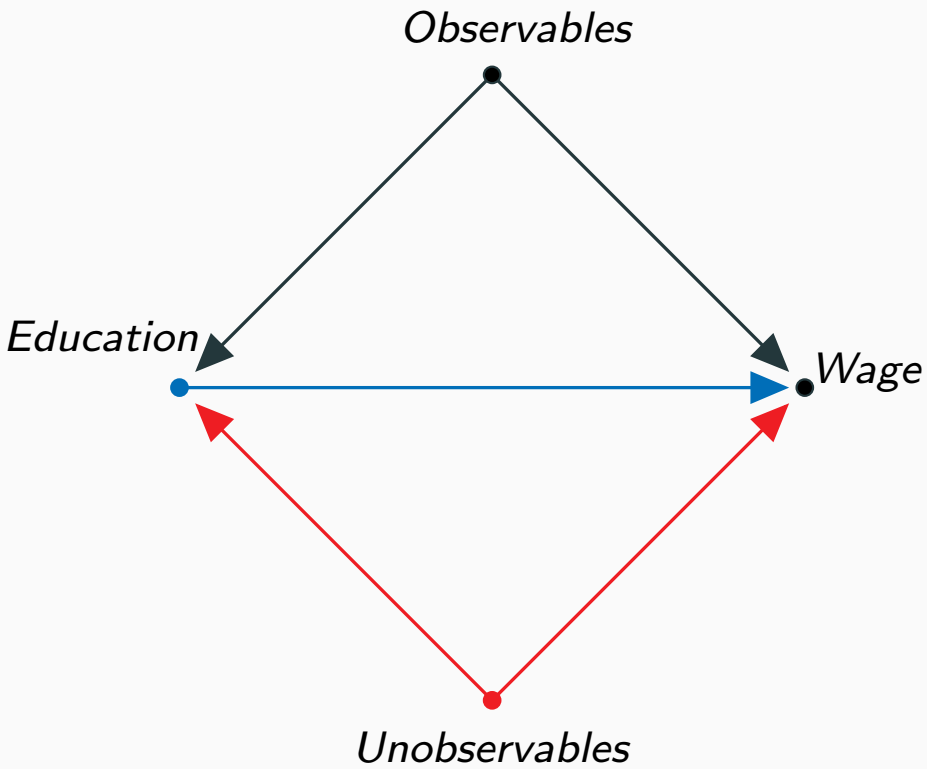
1. Draw a causal diagram representing it
2. What problems could you encounter in estimating the effect?

Causality and problems

Think about the causal claim you hypothesize.

1. Draw a causal diagram representing it
2. What problems could you encounter in estimating the effect?
3. How can you represent these problems?

Threats to causality



From problems to solutions

Think about what ways you could solve these problems.

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias?

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias? Why?

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias? Why?
2. What solutions do we have, with observational data, to replicate the conditions of an experiment and **identify** the causal effect?

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias? Why?
2. What solutions do we have, with observational data, to replicate the conditions of an experiment and **identify** the causal effect?
    - Instrumental variable

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias? Why?
2. What solutions do we have, with observational data, to replicate the conditions of an experiment and **identify** the causal effect?
    - Instrumental variable
    - Regression discontinuity

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias? Why?
2. What solutions do we have, with observational data, to replicate the conditions of an experiment and **identify** the causal effect?
    - Instrumental variable
    - Regression discontinuity
    - Difference-in-differences

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias? Why?
2. What solutions do we have, with observational data, to replicate the conditions of an experiment and **identify** the causal effect?
    - Instrumental variable
    - Regression discontinuity
    - Difference-in-differences
3. What **identification strategy** can you adopt, with these observational data?

From problems to solutions

Think about what ways you could solve these problems.

1. What ideal experiment would let you estimate the effect without bias? Why?
2. What solutions do we have, with observational data, to replicate the conditions of an experiment and **identify** the causal effect?
    - Instrumental variable
    - Regression discontinuity
    - Difference-in-differences
3. What **identification strategy** can you adopt, with these observational data?
4. What do you need to implement it?

Solutions

1. Controlling for observable does not eliminate the bias

Solutions

1. Controlling for observable does not eliminate the bias $\rightarrow$ unobservable are still out there!

Solutions

1. Controlling for observable does not eliminate the bias $\rightarrow$ unobservable are still out there!
2. In an ideal experiment we would manipulate the years of schooling and randomly assign them to respondents.

Solutions

1. Controlling for observable does not eliminate the bias $\rightarrow$ unobservable are still out there!
2. In an ideal experiment we would manipulate the years of schooling and randomly assign them to respondents. This way we would eliminate the effect of unobservable confounders on both education and wage

Solutions

1. Controlling for observable does not eliminate the bias $\rightarrow$ unobservable are still out there!
2. In an ideal experiment we would manipulate the years of schooling and randomly assign them to respondents. This way we would eliminate the effect of unobservable confounders on both education and wage
3. Given these observational data, we'd probably need to **instrument** our independent variable of interest (education)

Solutions

1. Controlling for observable does not eliminate the bias $\rightarrow$ unobservable are still out there!
2. In an ideal experiment we would manipulate the years of schooling and randomly assign them to respondents. This way we would eliminate the effect of unobservable confounders on both education and wage
3. Given these observational data, we'd probably need to **instrument** our independent variable of interest (education)
4. We need a variable that is affecting education but exogenous to our problem (does not impact wage other than through education)

Solutions

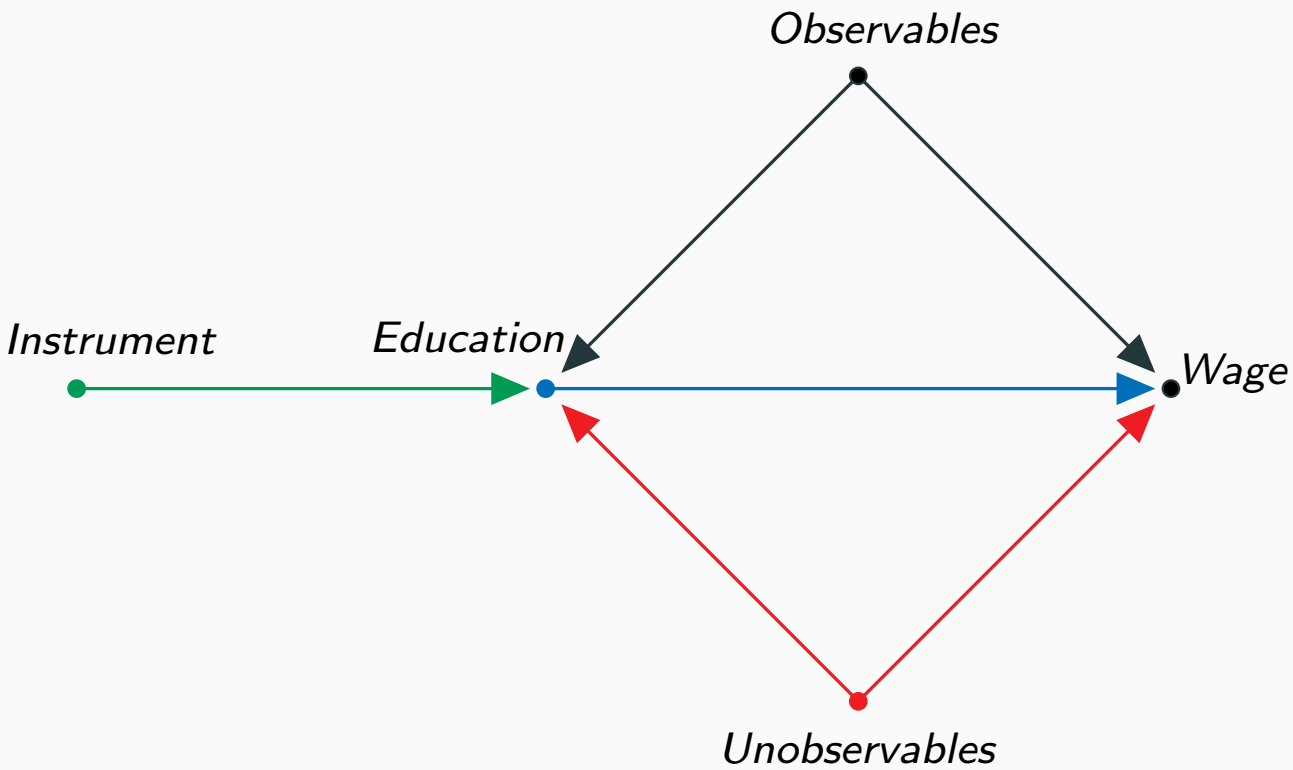
1. Controlling for observable does not eliminate the bias $\rightarrow$ unobservable are still out there!
2. In an ideal experiment we would manipulate the years of schooling and randomly assign them to respondents. This way we would eliminate the effect of unobservable confounders on both education and wage
3. Given these observational data, we'd probably need to **instrument** our independent variable of interest (education)
4. We need a variable that is affecting education but exogenous to our problem (does not impact wage other than through education)$\rightarrow$ exclusion restriction

Solutions

1. Controlling for observable does not eliminate the bias $\rightarrow$ unobservable are still out there!
2. In an ideal experiment we would manipulate the years of schooling and randomly assign them to respondents. This way we would eliminate the effect of unobservable confounders on both education and wage
3. Given these observational data, we'd probably need to **instrument** our independent variable of interest (education)
4. We need a variable that is affecting education but exogenous to our problem (does not impact wage other than through education)$\rightarrow$ exclusion restriction

What instruments could we use in general? Which variable in this dataset?

Solution: Instrumental variable!



Conduct your analysis entirely

Now, based on:

Conduct your analysis entirely

Now, based on:

1. Your research question

Conduct your analysis entirely

Now, based on:

1. Your research question
2. The problem you have for causal inference

Conduct your analysis entirely

Now, based on:

1. Your research question
2. The problem you have for causal inference
3. The identification strategy you have chosen

Conduct the analysis **entirely**:

Conduct your analysis entirely

Now, based on:

1. Your research question
2. The problem you have for causal inference
3. The identification strategy you have chosen

Conduct the analysis **entirely**:

1. Perform a descriptive statistical analysis of the variables of interest

Conduct your analysis entirely

Now, based on:

1. Your research question
2. The problem you have for causal inference
3. The identification strategy you have chosen

Conduct the analysis **entirely**:

1. Perform a descriptive statistical analysis of the variables of interest You can use all sort of plots (univariate but also bivariate or multivariate), summary statistics, etc.

Conduct your analysis entirely

Now, based on:

1. Your research question
2. The problem you have for causal inference
3. The identification strategy you have chosen

Conduct the analysis **entirely**:

1. Perform a descriptive statistical analysis of the variables of interest You can use all sort of plots (univariate but also bivariate or multivariate), summary statistics, etc.
2. Write down the equation(s) of your research design

Conduct your analysis entirely

Now, based on:

1. Your research question
2. The problem you have for causal inference
3. The identification strategy you have chosen

Conduct the analysis **entirely**:

1. Perform a descriptive statistical analysis of the variables of interest You can use all sort of plots (univariate but also bivariate or multivariate), summary statistics, etc.
2. Write down the equation(s) of your research design
3. Estimate the model using your preferred statistical software

Conduct your analysis entirely

Now, based on:

1. Your research question
2. The problem you have for causal inference
3. The identification strategy you have chosen

Conduct the analysis **entirely**:

1. Perform a descriptive statistical analysis of the variables of interest You can use all sort of plots (univariate but also bivariate or multivariate), summary statistics, etc.
2. Write down the equation(s) of your research design
3. Estimate the model using your preferred statistical software
4. Interpret the results and present them to the class

Conduct your analysis entirely

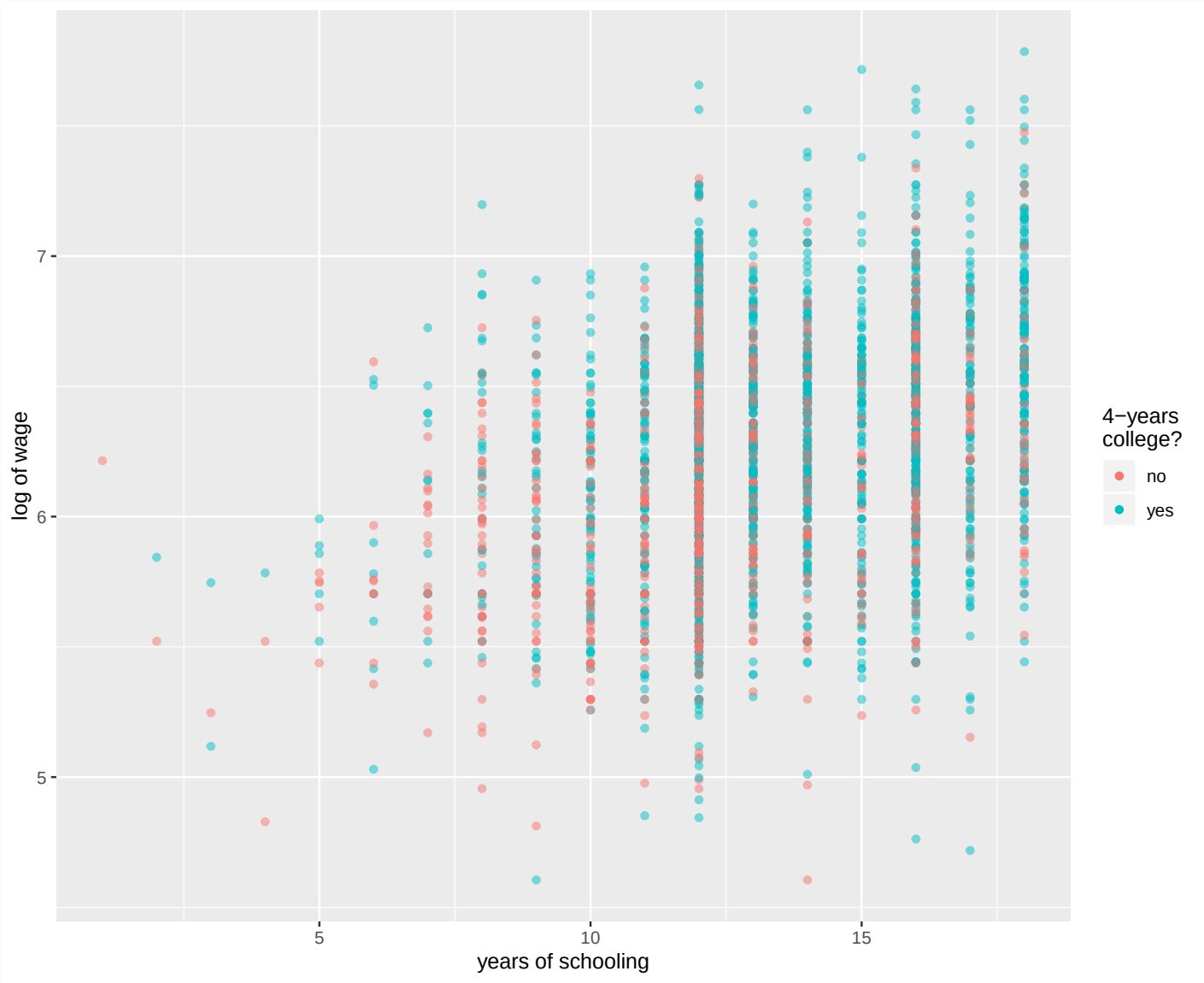
Now, based on:

1. Your research question
2. The problem you have for causal inference
3. The identification strategy you have chosen

Conduct the analysis **entirely**:

1. Perform a descriptive statistical analysis of the variables of interest You can use all sort of plots (univariate but also bivariate or multivariate), summary statistics, etc.
2. Write down the equation(s) of your research design
3. Estimate the model using your preferred statistical software
4. Interpret the results and present them to the class
5. Do you buy these results? What do you think of the identification strategy?

Descriptive statistics



Analysis

The author estimates the causal effect of $education_i$ on $log(wage_i)$ with a two-stage least squares (2SLS) procedure:

Analysis

The author estimates the causal effect of $education_i$ on $log(wage_i)$ with a two-stage least squares (2SLS) procedure:

$$\widehat{education}_i = \hat{\gamma} + \hat{\delta} college_i + \mathbf{X_i'}\hat{\eta}$$

Analysis

The author estimates the causal effect of $education_i$ on $log(wage_i)$ with a two-stage least squares (2SLS) procedure:

$$\widehat{education}_i = \hat{\gamma} + \hat{\delta} college_i + \mathbf{X_i'}\hat{\eta}$$

$$log(wage_i) = \hat{\alpha} + \hat{\beta}\widehat{education}_i + \mathbf{X_i'}\hat{\theta} + u_i$$

Results

	Model 1	Model 2
(Intercept)	5.06^{***}	4.16^{***}
	(0.07)	(0.84)
educ	0.07^{***}	0.12^{**}
	(0.00)	(0.05)
exper	0.03^{***}	0.06^{***}
	(0.00)	(0.02)
black1	-0.17^{***}	-0.12^{**}
	(0.02)	(0.05)
south1	-0.13^{***}	-0.11^{***}
	(0.02)	(0.02)
married	-0.04^{***}	-0.03^{***}
	(0.00)	(0.01)
smsa1	0.18^{***}	0.15^{***}
	(0.02)	(0.03)
R^2	0.31	0.25
Adj. R^2	0.30	0.25
Num. obs.	3003	3003

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Problems

Are the two requirements for an IV met in this example?

Problems

Are the two requirements for an IV met in this example?

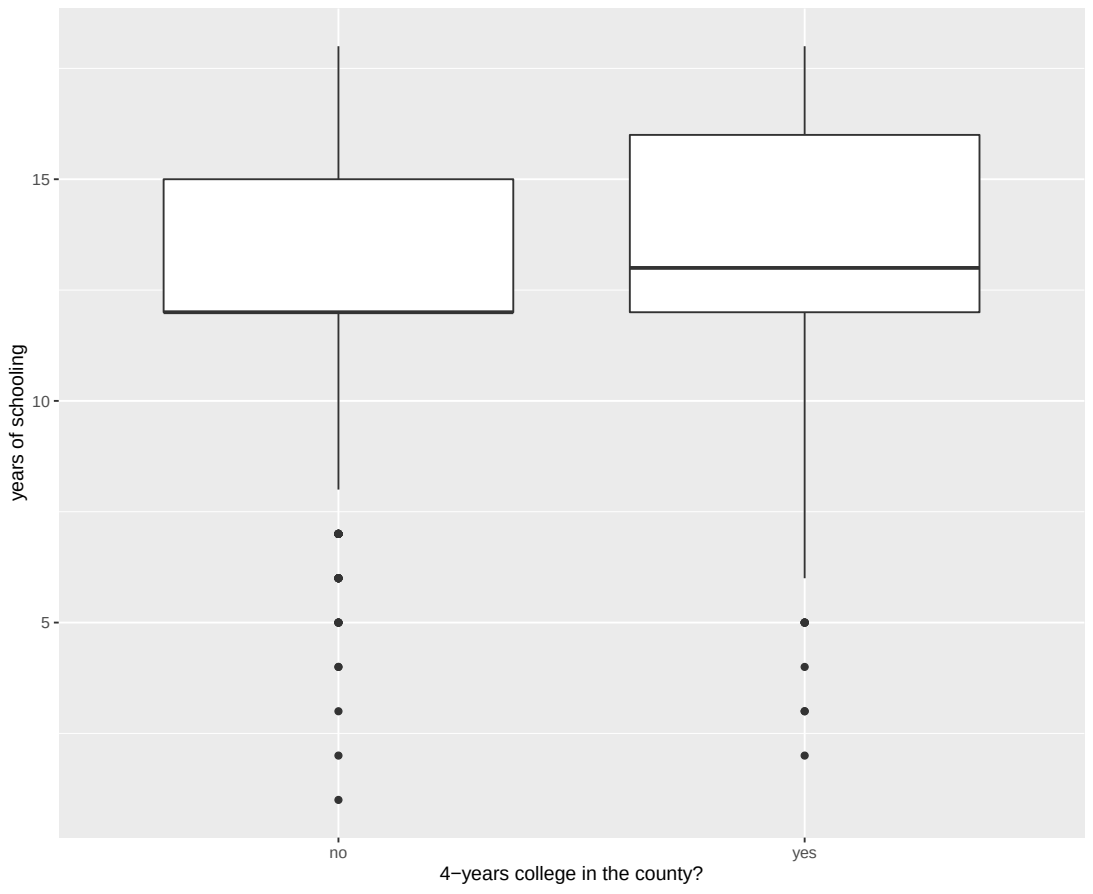
1. Strong instrument

Problems

Are the two requirements for an IV met in this example?

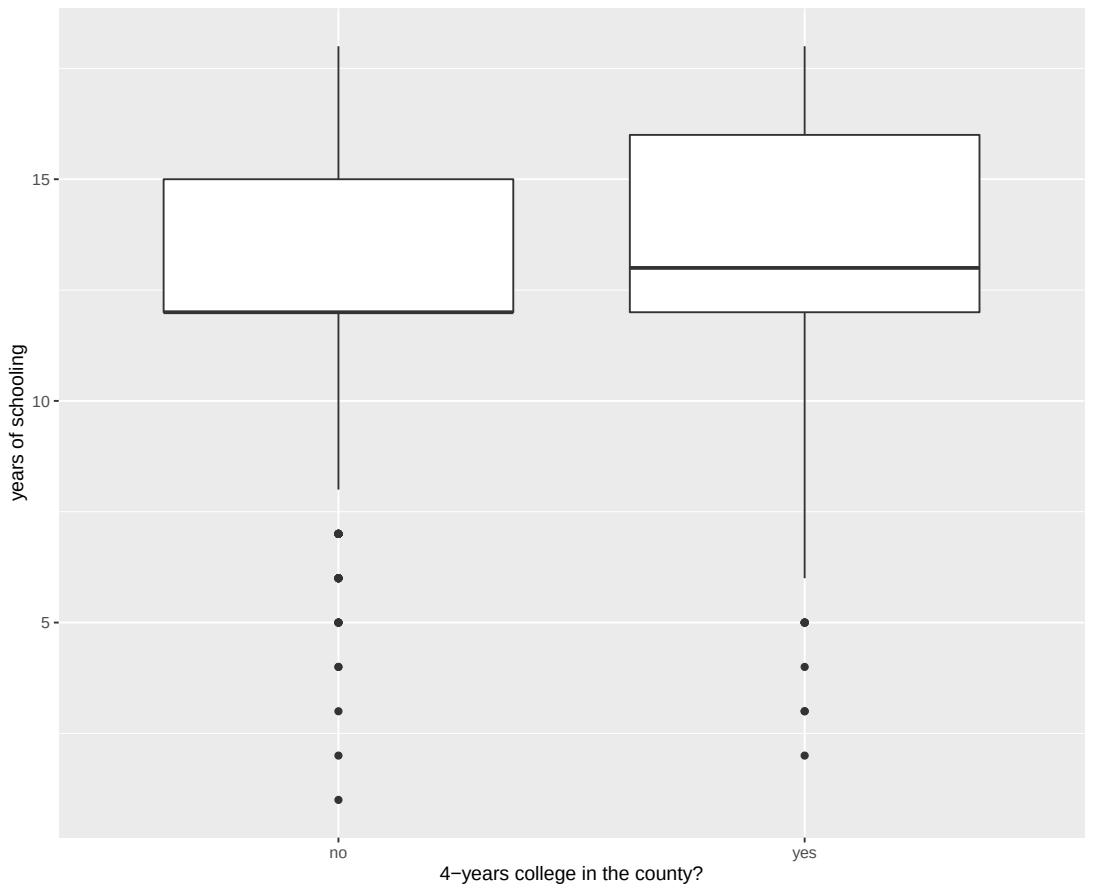
1. Strong instrument
2. Exclusion restriction

Strong instrument?



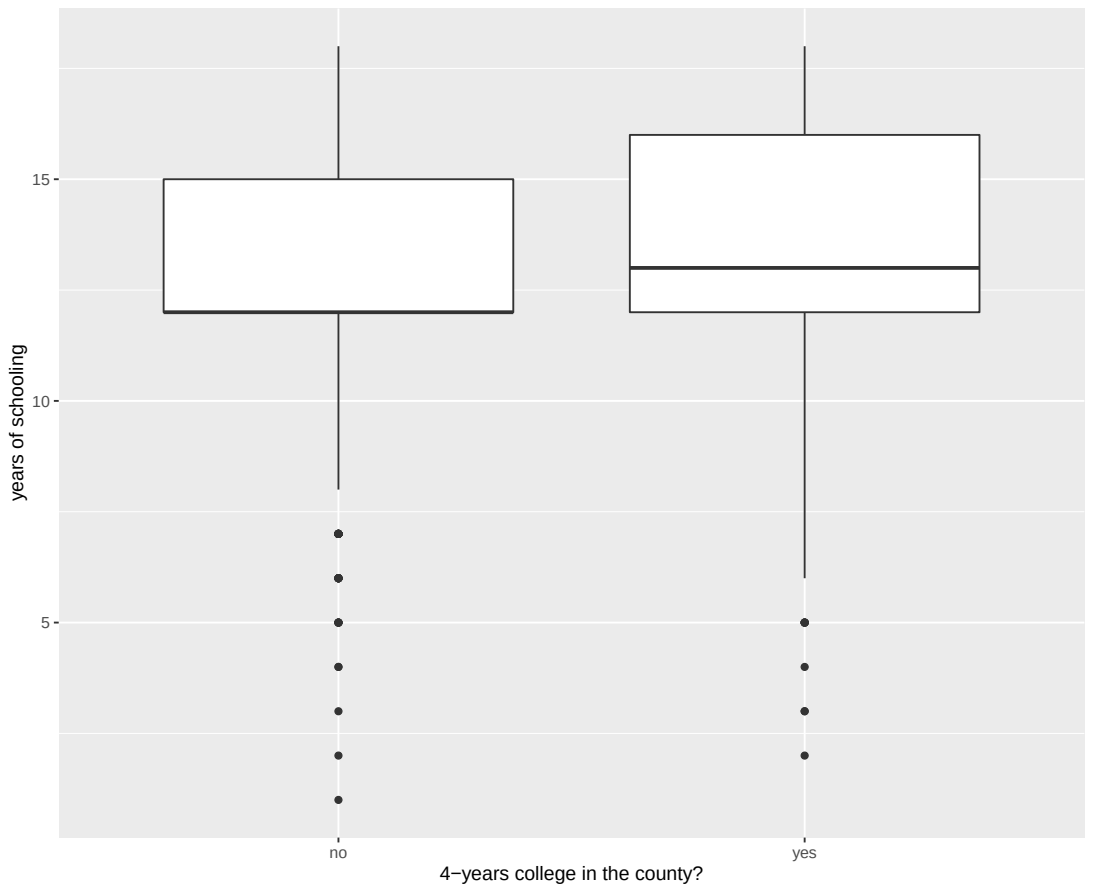
Does $college_i$ affect years of schooling?

Strong instrument?



Does $college_i$ affect years of schooling? Maybe after controlling for observables.

Strong instrument?



Does $college_i$ affect years of schooling? Maybe after controlling for observables.

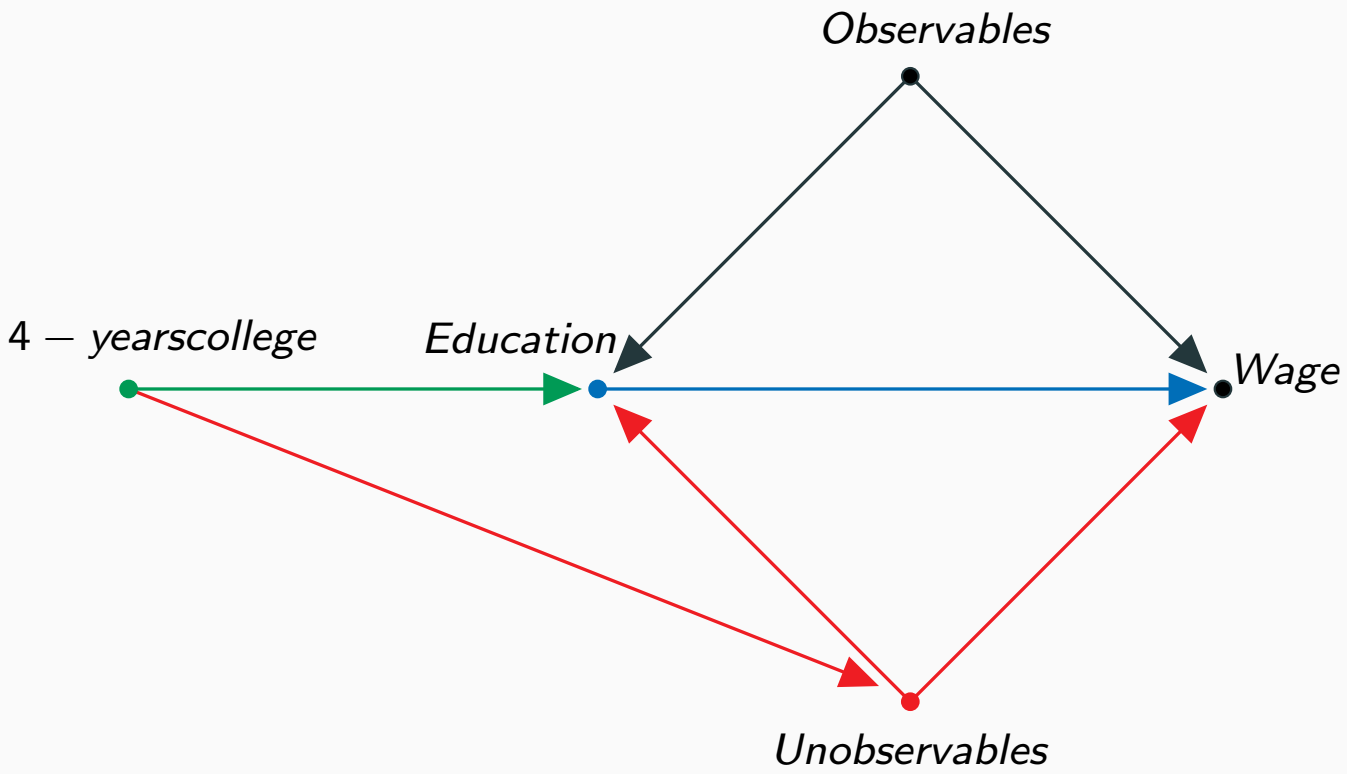
What about the exclusion restriction?

Exclusion restriction

Are we sure this never happens?

Exclusion restriction

Are we sure this never happens?



Exclusion restriction

What do you think about it? Can you think of any variable passing by which the presence of 4-years college affects wage?

Exclusion restriction

What do you think about it? Can you think of any variable passing by which the presence of 4-years college affects wage?

Examples:

Exclusion restriction

What do you think about it? Can you think of any variable passing by which the presence of 4-years college affects wage?

Examples:

- Social capital ($\neq$ *education*!)

Exclusion restriction

What do you think about it? Can you think of any variable passing by which the presence of 4-years college affects wage?

Examples:

- Social capital ($\neq$ *education*!)
- Income of a county

Exclusion restriction

What do you think about it? Can you think of any variable passing by which the presence of 4-years college affects wage?

Examples:

- Social capital ($\neq$ *education*!)
- Income of a county
- Possibilities of networking

Exclusion restriction

What do you think about it? Can you think of any variable passing by which the presence of 4-years college affects wage?

Examples:

- Social capital ($\neq$ *education*!)
- Income of a county
- Possibilities of networking

What could a good instrument be, instead?

Conclusion

All clear? More questions?

Thanks and see you next week!

References

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. *NBER working paper*, (w4483).