Announcements

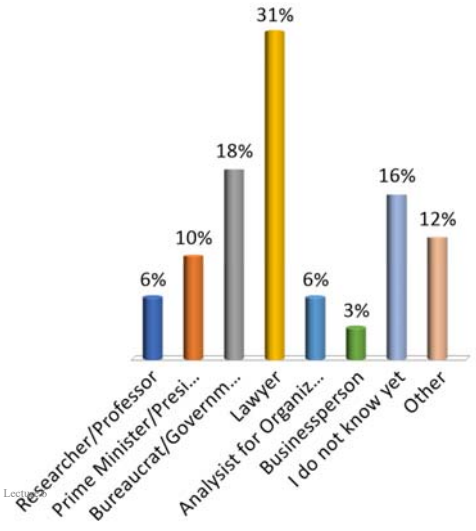
- Assignment 1, please finish by Saturday this week
- Assignment 2 until Sunday Sept 29, Questions?
- Conferences/Labs: R and R Studio download and check conference folder in **advance**, wor through Lab Guide **after** the conference
- Contact:
  - Conference and R related: your TA or Colin
  - Assignments: TA or Prof
  - Quiz related: Professor
  - Clarifications reg lectures: Professor
  - Missing lecture: is not encouraged, however see policy, no contact
- Midterm dates: Oct 17 and Nov 28
- Lecture time until 5.10pm from now on

Plan for today and Tuesday

- Common problems in causal inference
- Observational studies versus experiments
- How to best approach observational studies: Quasi-experimental designs
- Experimental designs

What would you like to be in the future?

1. Researcher/Professor
2. Prime Minister/President
3. Bureaucrat/Government Official
4. Lawyer
5. Analysist for Organization(s)
6. Businessperson
7. I do not know yet
8. Other



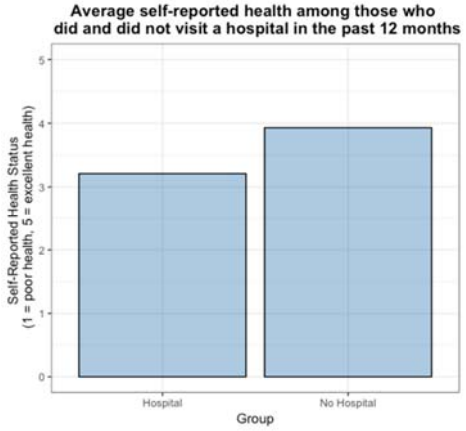
Do hospitals make people healthier?

Is this correct?

Individuals going to a hospital are *on average* less healthy than individuals not going to a hospital.

Survey respondents who reported going to a hospital in the past 12 months reported lower levels of health ($M = 3.21$) compared to those who did not go to a hospital ($M = 3.93$).

Conclusion: So going to the hospital makes people sick.



Source: National Health Interview Survey, 2005

Is that correct?

- Individuals going to the hospital are on average less healthy than individuals not going to the hospital;
- Individuals go to the hospital **after getting sick**; not before
- Individuals going to the hospital might be more likely to assess negatively their health compared to individuals not going to the hospital

Causal effects

Causal effect is the difference in potential outcome due to treatment (independent variable)

To answer causal questions, we must infer a **counterfactual** outcome and compare it with what actually happens (i.e., a factual outcome)

- E.g., What would happen if someone who went to the hospital decided to stay home? What if someone who had stayed home went to the hospital?

These considerations reflect different potential outcomes.

Potential Outcomes Framework

- **Key causal question**: Would the same individual give a different health score, had she not gone to the hospital?
- We can never observe both outcomes at the same time. This is the **fundamental problem of causal inference**; we observe only one of two potential outcomes.

Survey Respon-dent	Visited hospital?	Average self-reported health		Age	Highest Education Degree
		$Y_i(1)$	$Y_i(0)$		
1	1	2	?	55	BA
2	0	?	3	42	High School
3	0	?	4	19	BA
4	1	3	?	31	Some College
…	…	…	…	…	…
n	1	1	?	72	High school

Problem of causality

We are interested in testing for the effectiveness of a given treatment (e.g. tax incentives, good institutions, civic education courses)

If we just compare treated vs. non-treated individuals, we are typically not going to get the causal effect right

Why? Because individuals self-select into treatment, or treatment is offered to a select sample of individuals

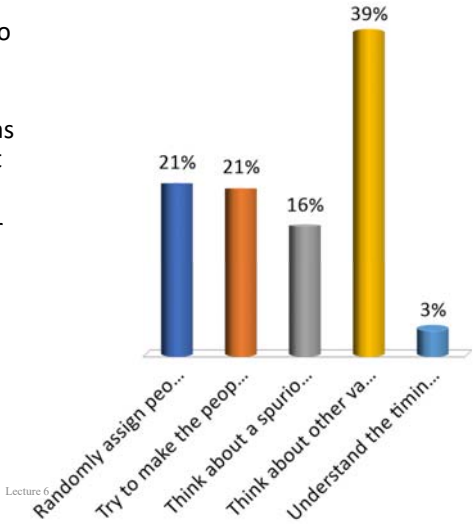
Whenever you read that X causes Y, ask yourself:

is there anything special about those with high values of X?

is that independent from Y?

How can we find out the causal story in the hospital example? Choose the best and most feasible answer.

1. Randomly assign people to go to the hospital or not
2. Try to make the people who go to the hospital as similar as possible to those who do not go
3. Think about a spurious factor
4. Think about other variables that explain differences in health scores
5. Understand the timing of getting sick



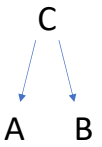
Examples of Causal Inference Problems

Observed relationships between A and B

Assumption/Theory: A → B

BUT: Reverse causality: A ← B

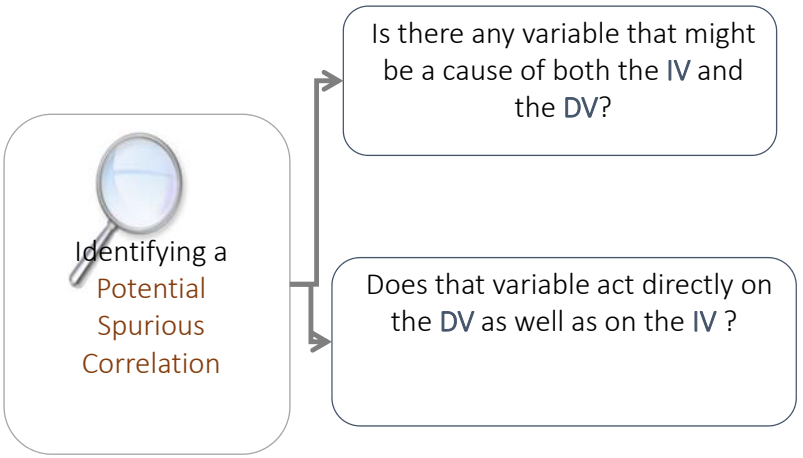
Spurious relationship:



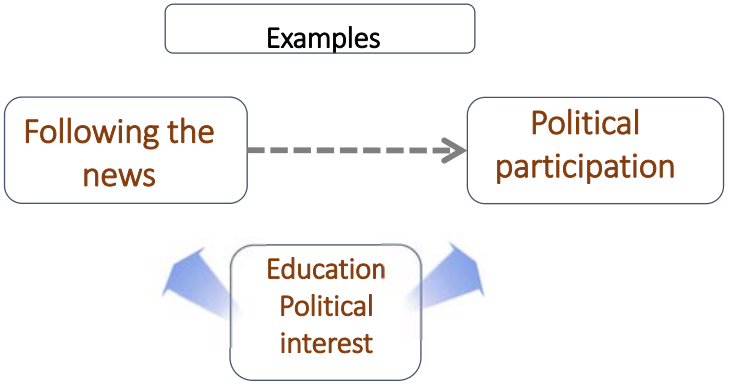
Or there can be no relationship at all

Selection effects are common causes of spurious relationships. Possible examples: "gateway drugs," member-ship in associations, hospital stays.

Identifying Sources of Spuriousness

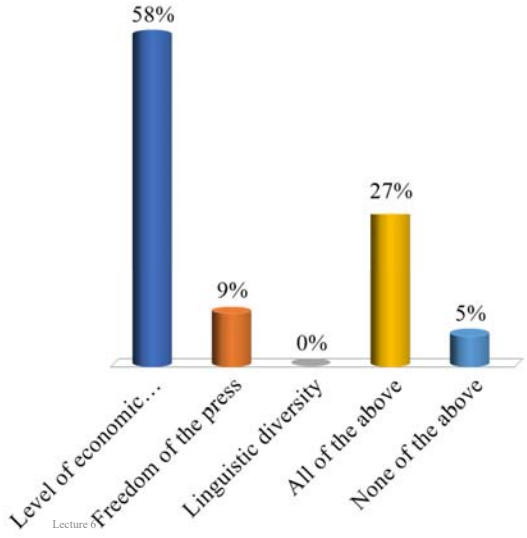


Identifying Sources of Spuriousness

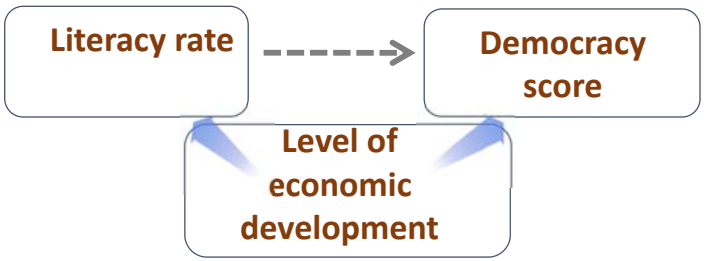


"The higher a country's literacy rate, the more democratic it will be." A plausible source of spuriousness is:

1. Level of economic development
2. Freedom of the press
3. Linguistic diversity
4. All of the above
5. None of the above



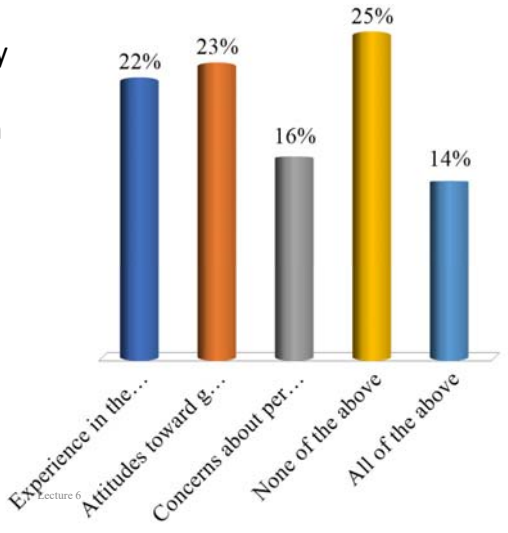
The higher a country's literacy rate, the more democratic it will be.



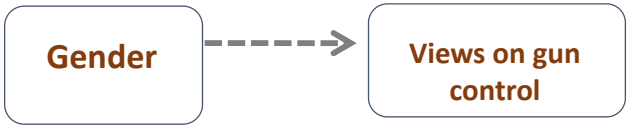
Freedom of the press would be an indicator of how democratic a country is, not a cause of how literate the population is or how democratic the country is

"Men are more likely than women to oppose gun control." A plausible source of spuriousness is:

1. Experience in the military
2. Attitudes toward government intervention
3. Concerns about personal safety
4. None of the above
5. All of the above

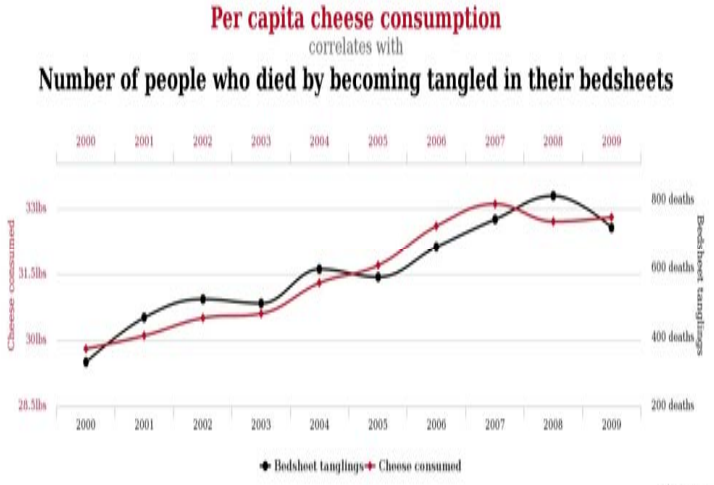


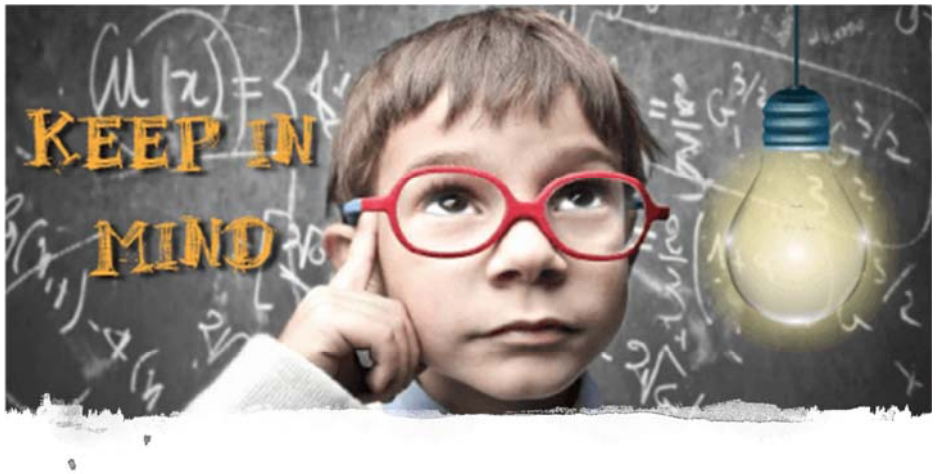
Men are more likely than women to oppose gun control



There is NO plausible source of spuriousness since nothing that could be measured in a survey would "cause" a person's gender

Examples of Causal Inference Problems





Keep in mind!

- Common problems in causal inference:
  - Spuriousness
  - Reverse causality

Why is an experimental design so advantageous for addressing causality?

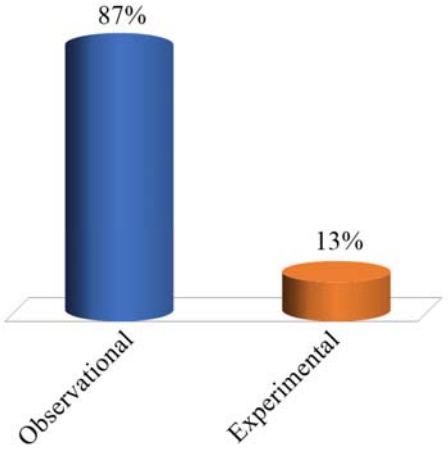
- Time Order
- Other influences/factors can be excluded
- Random assignment prevents self-selection into treatment (control group)
- However: internal vs. external validity

Observational Study vs. Experiment

- In an *observational study,* the researcher observes values of the dependent and independent variables for the sampled subjects (cases), without anything being done to the subjects (such as imposing a treatment)
- A researcher conducts a "true" *experiment* by *assigning* subjects randomly to certain experimental conditions and then observing outcomes (DV). The experimental conditions are called *treatments* and represent the IV of the study.

The researchers find 100 men aged 30 of which 50 have been smoking a pack a day for 10 years while the other 50 have been smoke free for 10 years. They then measure lung capacity for each of the 100 men. What kind of study is this?

1. Observational
2. Experimental



Experimental counterpart

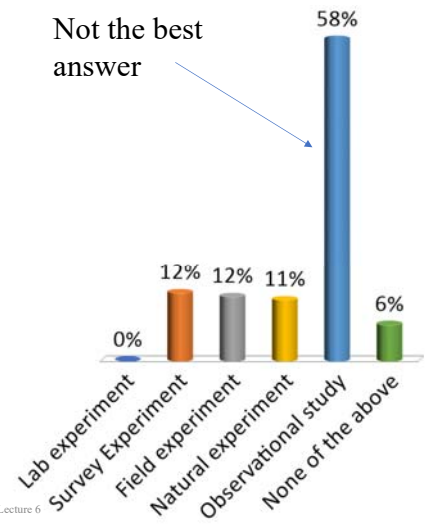
- The researcher finds 100 women aged 20 who do not currently smoke.
- They **randomly assign** 50 of the 100 women to the smoking treatment and the other 50 to the non-smoking treatment.
- Those in the smoking group smoke a pack a day for 10 years while those in the control group remain smoke free for 10 years.
- They then measure lung capacity for each of the 100 women.

Addressing the Causality Issue in Empirical Research

- Using the experimental logic for observational studies
  - Quasi-Experiments
- Using a variety of experimental designs, not just classic experiments
  - Natural Experiments
  - Lab Experiments
  - Field Experiments
  - Survey Experiments

What kind of study did Eitan Hersh conduct?

1. Lab experiment
2. Survey Experiment
3. Field experiment
4. Natural experiment
5. Observational study
6. None of the above



Conclusion

- Use quasi-experimental design logic whenever you can
- Tuesday we look at a variety of experimental designs
- Read or re-read Eitan Hersh, Ryan Enos, and ch 11 in textbook
- Reading quiz Tuesday
- Class Survey (assignment 1)

Team Scores

Points	Team	Points	Team
2.4	**Businessperson**		
2.21	**I do not know yet**		
2.06	**Bureaucrat/Governmen...**		
2.06	**Lawyer**		
2.01	**Analysist for Organi...**		
1.97	**Prime Minister/President**		
1.9	**Researcher/Professor**		