SOC 3811/5811:
BASIC SOCIAL STATISTICS

Causal Inference and Observational Data

Causality and Counterfactuals

When we talk about cause and effect in the social sciences, we usually talk about how one variable affects another

The independent variable is said to affect or cause change in the dependent variable

Example: "Does poverty cause kids to commit crimes?"

Independent Variable: Whether a child is poor

Dependent Variable: Whether a child commits crimes

Causality and Counterfactuals

Independent Variable	Dependent Variable
Explanatory Variable	Response Variable
Predictor Variable	Outcome Variable
Upstream Variable	Downstream Variable
Exogenous Variable	Endogenous Variable
Regressor Variable	Regressand Variable
Cause	Effect

Example #1

UNIVERSITY OF MICHIGAN

NEWS SERVICE

OFFICE OF THE VICE PRESIDENT FOR GLOBAL COMMUNICATIO

HOME | NEWS RELEASES | MULTIMEDIA | CONTACTS

Air pollution near Michigan schools linked to poorer student health, academic performance

Published on May 20, 2011
Contact Jim Erickson

0

ANN ARBOR, Mich.—Air pollution from industrial sources near Michigan public schools jeopardizes children's health and academic success, according to a new study from University of Michigan researchers.

The researchers found that schools located in areas with the state's highest industrial air pollution levels had the lowest attendance rates (an indicator of poor health) as well as the highest proportions of students who failed to meet state educational testing standards. The researchers examined the distribution of all 3,660 public elementary, middle, junior high and high schools in the state and found that 62.5 percent of them were located in places with high levels of air pollution from industrial sources. Minority students appear to bear the greatest burden, according to a research team led by Paul Mohai of the U-M School of Natural Resources and Environment and Byoung-Suk Kweon of the U-M Institute for Social Research.

http://ns.umich.edu/new/releases/8398

Example #2

Breast implants linked with suicide in study

By Maggie Fox, Health and Science Editor

WASHINGTON | Wed Aug 8, 2007 7:10pm EDT

Related Topics

Health »
Lifestyle »

(Reuters) - Women who get cosmetic breast implants are nearly three times as likely to commit suicide as other women, U.S. researchers reported on Wednesday.

The study, published in the Annals of Plastic Surgery, reinforces several others that have shown women who have breast enlargements have higher suicide risks.

Loren Lipworth of the Vanderbilt University Medical Center in Tennessee and colleagues followed up on 3,527 Swedish women who had cosmetic breast implant surgery between 1965 and 1993. They looked at death certificates to analyze causes of death among women with breast implants.

Only 24 of the women had committed suicide after an average of 19 years, but this worked out to triple the risk compared to the average population, they reported. Doctors who perform cosmetic breast surgery may want to monitor patients closely or screen them for suicide risk, Lipworth said.

http://www.reuters.com/article/2007/08/08/us-implants-suicide-idUSN08369190200708 08?feedType=RSS&rpc=22&sp=true

Example #3

Vegetarians are more intelligent, says study

Recommend 59 Tweet 0 Share +1 0

15 December 2006

Frequently dismissed as cranks, their fussy eating habits tend to make them unpopular with dinner party hosts and guests alike.

But now it seems they may have the last laugh, with research showing vegetarians are more intelligent than their meat-eating friends.

A study of thousands of men and women revealed that those who stick to a vegetarian diet have IQs that are around five points higher than those who regularly eat meat.

Writing in the British Medical Journal, the researchers say it isn't clear why veggies are brainier - but admit the fruit and veg-rich vegetarian diet could somehow boost brain power.

The researchers, from the University of Southampton, tracked the fortunes of more than 8,000 volunteers for 20 years.

Share

Related Articles

Hot new restaurants opening in November

Patient recovers after stroke kills half his brain

Billie Piper and husband Laurence Fox have to rehome their good life livestock

Suggested Topics

The Brain

Vegetarian

Top stories in News

Freddie Star arrested in Jimmy Savile inquiry

Hurricane Sandy: benefit concert will see Bruce Springsteen, Bon Jovi and Billy Joel take to the stage

Hurricane Sandy: undamaged parts of New York's subway reopen

http://www.standard.co.uk/news/vegetarians-are-more-intelligent-says-study-7082629.html

Example #4

The Effect of Country Music on Suicide*

Steven Stack
Direct correspondence to Steven Stack, Department of Sociology, Wayne State University, Detroit, MI 48202.

Jim Gundlach
+ Author Affiliations

Abstract

This article assesses the link between country music and metropolitan suicide rates. Country music is hypothesized to nurture a suicidal mood through its concerns with problems common in the suicidal population, such as marital discord, alcohol abuse, and alienation from work. The results of a multiple regression analysis of 49 metropolitan areas show that the greater the airtime devoted to country music, the greater the white suicide rate. The effect is independent of divorce, southernness, poverty, and gun availability. The existence of a country music subculture is thought to reinforce the link between country music and suicide. Our model explains 51% of the variance in urban white suicide rates.

http://sf.oxfordjournals.org/content/71/1/211.short

Example #5

Guns in Homes Strongly Associated with Higher Rates of Suicide

Suicidal Acts Using Firearms Highly Lethal Compared to Other Means

For immediate release: Tuesday, April 10, 2007

Boston, MA -- In the first nationally representative study to examine the relationship between survey measures of household firearm ownership and state level rates of suicide in the U.S., researchers at the Harvard School of Public Health (HSPH) found that suicide rates among children, women and men of all ages are higher in states where more households have guns. The study appears in the April 2007 issue of *The Journal of Trauma.*

"We found that where there are more guns, there are more suicides," said Matthew Miller, Assistant Professor of Health Policy and Management at HSPH and lead author of the study.

Suicide ranks as one of the 15 leading causes of death in the U.S.; among persons less than 30 years old, it is one of the top three causes of death. In 2004, more than half of the 32,439 Americans who committed suicide used a firearm.

http://www.hsph.harvard.edu/news/press-releases/2007-releases/press04102007.html

Counterfactuals

Causality and Counterfactuals

At the heart of all case and effect statements is a *counterfactual—the situation that would have existed had the explanatory variable not changed*

For example, by making the causal claim:

> "Living in poverty caused kids to commit crimes"

we are simultaneously making the counterfactual claim:

> "If the kids had **not** lived in poverty, then they would **not** have committed crimes"

A fundamental problem in explanatory research is that we never actually observe the counterfactual

Worksheet

What are the counterfactual claims that are implied by these causal statements?

1. Air pollution affects kids' school performance
2. Breast implants affect women's suicide rates
3. Eating a vegetarian diet makes you smarter
4. Listening to country music leads to suicide
5. Guns in homes lead to more suicides

Causality and Counterfactuals

The fact that we do not observe the counterfactual presents a major logical problem

How do we know what the value of the response variable would have been had the explanatory variable taken on a different value?

We **cannot** get around this fundamental logical problem …

…but we **can** design research in such a way that we can make good inferences about the nature of the counterfactual

Causality and Counterfactuals

Experimental research designs are—for reasons we'll cover—much stronger with respect to their ability to yield valid causal claims

However, we are typically only able to carry out non-experimental or observational research

Multivariate (multiple-variable) statistics are valuable for use in non-experimental, observational research because they help us *approximate* counterfactual situations

Criteria for Establishing Causality

Criteria for Establishing Causality

Three conditions that must be met in order to establish that X causes Y:

1. X and Y must be empirically associated (criteria of association)
2. Change in X must precede change in Y in time (criteria of temporal ordering)
3. There must be no third variable, Z, which acts as a "confounder"—or which induces "spuriousness"—in the association between X and Y (criteria of nonspuriousness)

Different research designs have their own ways of meeting these criteria, of facilitating inferences about the nature of the counterfactual, and thus of allowing us to make defensible causal claims

Criteria for Establishing Causality

Criteria of Association

In order to establish that X causes Y—or that change in the explanatory variable X causally leads to change in the response variable Y—we must observe an empirical association between X and Y

We have spent the past several weeks learning how to measure the association between variables (with our technique for doing so dependent upon whether X and Y are discrete, continuous, or a combination of the two)

Criteria for Establishing Causality

How would the association between X and Y be established in these examples?

1. Air pollution affects kids' school performance
2. Breast implants affect women's suicide rates
3. Eating a vegetarian diet makes you smarter
4. Listening to country music leads to suicide
5. Guns in homes lead to more suicides

Criteria for Establishing Causality

Criteria of Temporal Ordering

In order to establish that X causes Y—or that change in the explanatory variable X causally leads to change in the response variable Y—it must be the case that the change in X occurred before the change in Y

As we will discuss, this can be difficult in observational research

Criteria for Establishing Causality

Is the criteria of causal ordering clearly met in these examples? Does it depend on how you collected the data?

1. Air pollution affects kids' school performance
2. Breast implants affect women's suicide rates
3. Eating a vegetarian diet makes you smarter
4. Listening to country music leads to suicide
5. Guns in homes lead to more suicides

Criteria for Establishing Causality

Criteria of Nonspuriousness

Suppose we observe that shoe size at age 18 is associated with frequency of criminal behavior in adulthood, such that people with bigger shoes at age 18 commit more crimes later on

We've met the first two criteria required to make causal statement— but can we say that shoe size at age 18 causally affects criminal behavior in adulthood?

Criteria for Establishing Causality

Criteria of Nonspuriousness

If some third variable(s) affects both shoe size at age 18 and criminal behavior in adulthood, then the association between shoe size and criminal behavior is a spurious relationship, not a causal relationship

Can you think of any?

The “third variables” are called confounding variables

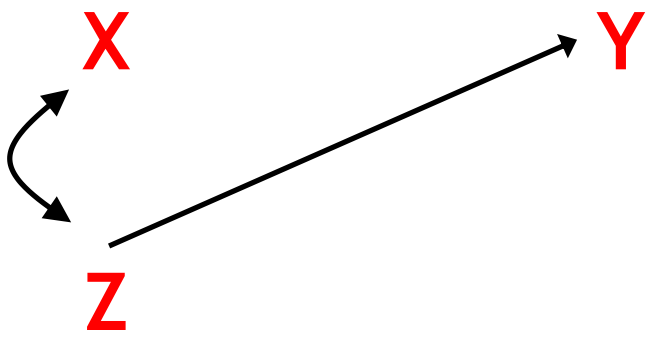
Criteria for Establishing Causality

Criteria of Nonspuriousness

What is the effect of X on Y?

In this example, X and Y are associated, but the association is *entirely* spurious owning to Z

Z is a counfounder



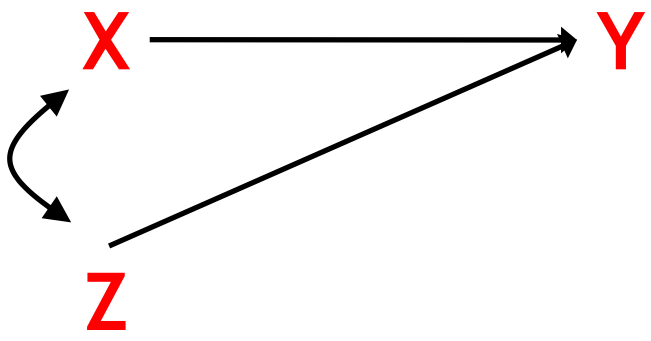
Criteria for Establishing Causality

Criteria of Nonspuriousness

What is the effect of X on Y?

Here X and Y are associated

That association is *partly* spurious (owing to confounder Z) and *partly* causal



Criteria for Establishing Causality

Could spuriousness threaten the validity of these causal claim?

1. Air pollution affects kids' school performance
2. Breast implants affect women's suicide rates
3. Eating a vegetarian diet makes you smarter
4. Listening to country music leads to suicide
5. Guns in homes lead to more suicides

Worksheet

Why might the association between A and B be spurious (or at least partly spurious) in the following examples? That is, why might the association between A and B not be entirely causal in nature?

1. Having a TV set in the bedroom (A) is correlated with couples having lower frequency of sexual activity (B) (research finding).
2. Listening to sexually explicit music lyrics (A) is correlated with teen sexual activity (B) (research finding).
3. Alcohol consumption (A) is correlated with violent behavior (B) (research findings).

Causality and Research Design

Causality and Research Design

In experimental research cases are randomly assigned to two or more comparison groups … X is defined by the group to which cases are assigned

The response variable Y is measured before (pre-test) and after (post-test) the manipulation of X

If the change in Y between the pre-test and the post-test differs across levels of X, then X and Y are associated

Because the value of X is assigned at random, spuriousness is not possible

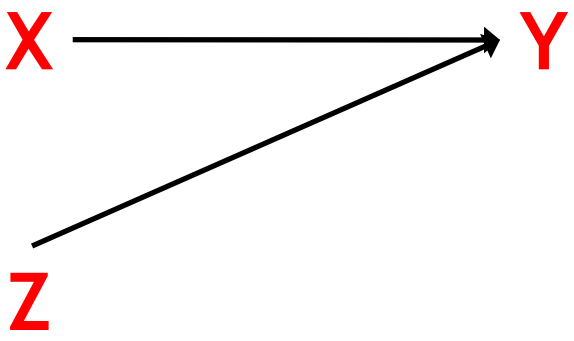
Causality and Research Design

Experimental Design

What is the effect of X on Y?

X and Y are associated; X precedes Y in time

Because the value of X is assigned at random, spuriousness is not possible



Causality and Research Design

Main weaknesses of experimental methods:

1. Experiments can be expensive or unethical to conduct in many circumstances
2. Experimental research does not typically yield any information about causal mechanisms
3. Randomly assigning cases to comparison groups is not the same thing as randomly selecting a sample … results from experiments are not always generalizable

Causality and Research Design

Observational research involves studying naturally-occurring variation in X (and Y), with no intervention from the researcher

There are several techniques for assessing the magnitude of association between X and Y

However, in observational research…

- …it can be difficult to establish temporal ordering
- …it is usually extremely difficult to rule out the possibility that the observed relationship is spurious

Causality and Research Design

Despite these threats to making valid causal statements, most social science research is observational

If we want to study the effects of most important social it is usually not possible (or ethical) to randomly assign cases to treatment and control groups

To establish temporal ordering, researchers sometimes use longitudinal (as opposed to cross-sectional) designs

To avoid problems with spuriousness, researchers attempt to statistically control for confounding variables

Causality and Research Design

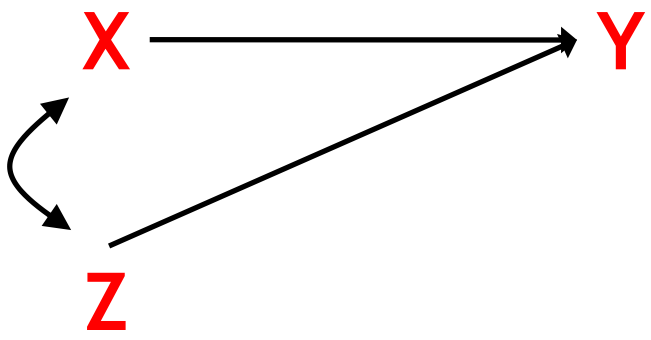
Observational Design

What is the effect of X on Y?

X and Y are associated; X may precede Y in time

The association between X and Y that remains after statistically controlling for Z is causal in nature

(but only if we statistically adjust or "control" for all Z variables!)



Statistical Control

Statistical Control

Statistical control is a set of technique used in observational research to reduce the risk of spuriousness

The confounding variables Z are “held constant” so that we can observe the association between X and Y “net of” Z

Conceptually, holding Z “constant” means observing the association between X and Y among people with equal values on the confounding Z variables

We presume that the association between X and Y that persists after statistically controlling for Z is causal in nature (as long as the other criteria have been met)

Statistical Control

Example: What is the effect of shoe size on criminality?

		Shoe Size at age 15		
		Small	Medium	Large
Criminal?	Yes	16	24	32
as an Adult?	No	3,984	3,976	3,968

Gamma = 0.22

Criteria of Association? … OK

Temporal Ordering Criteria? … OK

Nonspuriousness? … What about gender?

Statistical Control

Example: What is the effect of shoe size on criminality?

MEN

		Shoe Size at age 15		
		Small	**Medium**	**Large**
Criminal?	**Yes**	10	20	30
as an Adult?	**No**	990	1,980	2,970

Gamma = 0.00

WOMEN

		Shoe Size at age 15		
		Small	**Medium**	**Large**
Criminal?	**Yes**	6	4	2
as an Adult?	**No**	2,994	1,996	998

Gamma = 0.00

Statistical Control

Example: What is the effect of education on family income as an adult?

		Family Income as Adult		
		< Avg	Avg	> Avg
Education	< H.S.	2,141	2,959	440
	H.S.	2,346	5,252	1,403
	> H.S.	2,525	5,682	4,406

Source: GSS

Gamma = 0.37

Criteria of Association? … OK

Temporal Ordering Criteria? … OK

Nonspuriousness? … What about family background?

Statistical Control

Example: What is the effect of education on family income as an adult *net of father's education*?

Father: <H.S.

Education	Family Income as Adult: < Avg	Avg	> Avg
< H.S.	1,811	2,445	341
H.S.	1,461	3,106	731
> H.S.	908	1,981	1,300

Source: GSS

Gamma = 0.34

Father: H.S.

Education	Family Income as Adult: < Avg	Avg	> Avg
< H.S.	242	359	53
H.S.	613	1,594	467
> H.S.	684	1,776	1,190

Source: GSS

Gamma = 0.30

Father: >H.S.

Education	Family Income as Adult: < Avg	Avg	> Avg
< H.S.	88	155	46
H.S.	272	552	205
> H.S.	933	1,925	1,916

Source: GSS

Gamma = 0.32

The association between X and Y is partially spurious

Myths about Causality

Causality: Myths Dispelled

Myth #1: "X isn't the only (or even the most important) cause of Y, so it's wrong to say that X causes Y"

- Mechanical vs. probabilistic explanations
- Whether you get lung cancer depends on a lot of things … only one of which is cigarette smoking
- Just because smoking isn't the only factor that affects whether you get lung cancer doesn't mean it isn't a causal factor

Causality: Myths Dispelled

Myth #2: “Exceptions Disprove the Rule”

- Mechanical vs. probabilistic explanations
- Remember that causal statements pertain to how one *variable* affects another
- Even if smoking does cause lung cancer, we should still expect to find many smokers who don’t get lung cancer and many people with lung cancer who never smoked

Causality: Myths Dispelled

Myth #3: “X doesn’t always—or even usually—lead to Y, so therefore X doesn’t cause Y”

- Does driving drunk cause people to crash their cars?
- The vast majority of drunk drivers get home safely
- Does this mean that drunk driving has no effect on the odds of crashing? No…
- As long as the probability of crashing is higher for drunk drivers than for sober drivers — even if the probability is still very low for drunk drivers — then it is still true that driving drunk is causally related to crashing