Eat less salt, drink more wine, dump the cellphone, eat more salt, and live longer:

teaching students to understand the role of data collection in statistical inference

The message

- Statistical inference is concerned with answering the question "Could observed differences be simply due to chance?"
- To answer, we must choose an appropriate test or inferential procedure, check assumptions, calculate (or have calculated) appropriate test statistics, and compare results to an appropriate chance model.

The message

- This is familiar ground, because a background in mathematics prepares us well.
- But "is this due to chance" is only half the battle. The more interesting half is "What does this tell us about the world?"

The message

- And to answer this, we need to understand that data are not just numbers but, as statistician David Moore said:
- Data are numbers with context.
- And by "context" we usually mean the context in which the data were gathered.

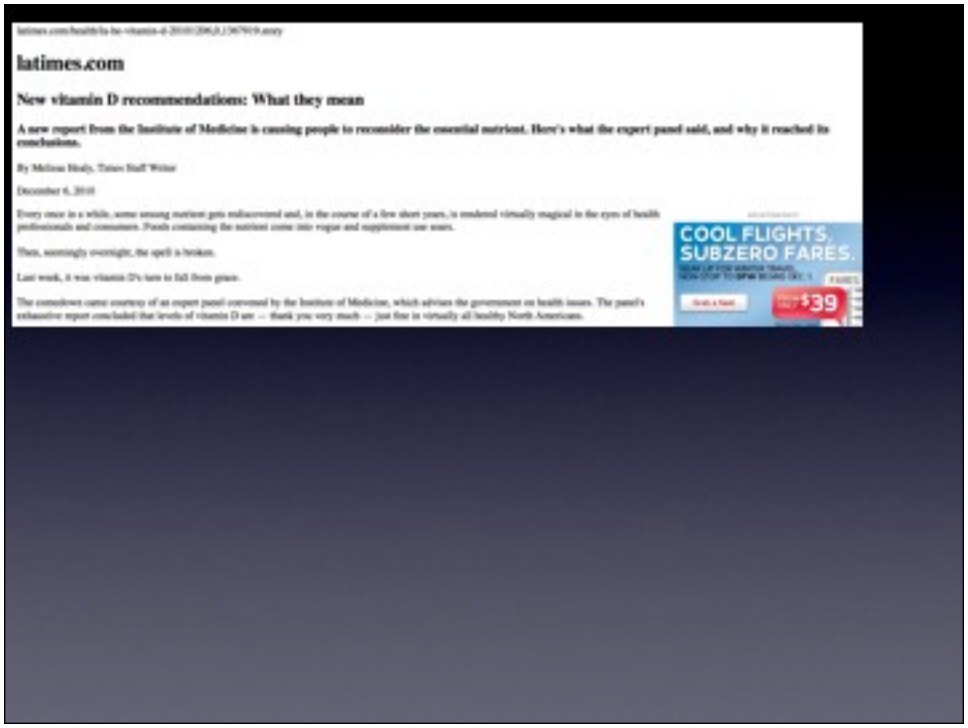
data collection and causal inference

- Statistical inference is limited or enhanced by the design of the data collection procedure.
- An important part of statistical inference is causal inference: X and Y might be associated, but can we also conclude that changing X *causes* a change in Y?

read "Video Games"

- Weaknesses of this study?
- Strengths of this study?
- What are the conclusions? Do they "hold up"?
- (Note: can you separate the conclusion of the news article from the conclusion of the study?)


Los Angeles Times
HEALTH
LOCAL U.S. WORLD BUSINESS SPORTS ENTERTAINMENT HEALTH LIVING TRAVEL OPINION MORE
BOOSTER SHOTS FITNESS & NUTRITION MEDICINE MENTAL HEALTH HEALTHCARE REFORM BREAST CANCER DIABETES
HEALTH TOPICS: PSYCHOLOGY FADS | NEW CHOLESTEROL DRUG | MEDICAL CARE QUALITY | HEALTHKEY: HEALTH A-Z | VINEGA
Drug shows promise in boosting 'good' cholesterol
A clinical study of the new drug anacetrapib shows a dramatic increase in HDL cholesterol levels, seen as a key to lowering heart disease risk, without the side effects seen in a previous drug.
32
14
0
By Shari Roan, Los Angeles Times
E-mail Print Share Text Size
32 people like this. Be the first of your friend


latimes.com
New vitamin D recommendations: What they mean
A new report from the Institute of Medicine is causing people to reconsider the essential nutrient. Here's what the expert panel said, and why it reached its conclusions.
By Melissa Healy, Times Staff Writer
December 6, 2010
COOL FLIGHTS. SUBZERO FARES.
$39

The Healthy Skeptic: Homeopathic cold remedies

Products may be too diluted to be helpful, or their active ingredients may not help much.

Zicam RapidMelts employs a dilution of a zinc compound. (Zicam. / September 25, 2006)

Others:

- cellphones causes brain cancer
- sleep deprivation impedes memory
- (moderate) alcohol consumption prolongs life
- humans affect climate change
- abortion lowers crime rates
- HIV causes AIDS

What is causality?

- If a change in the value of x *tends to* result in a change in the value of y, then we say x causes y.
- For example, if I change my status from "no flu shot" to "flu shot", then I am less likely to get the flu.
- Note that I could still get the flu, without invalidating the effectiveness of the vaccine.

paradigm

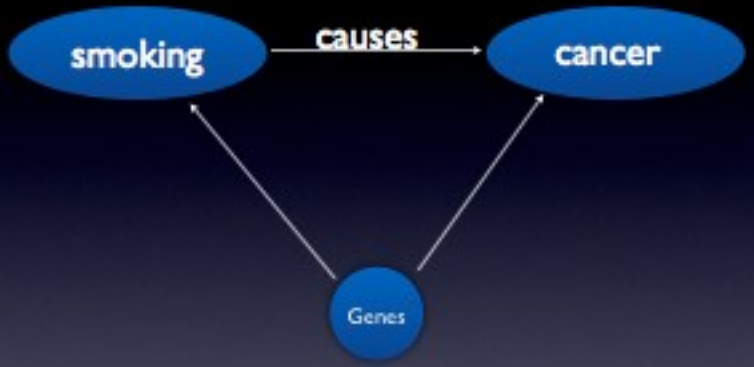
- Two groups: Treatment and Control. Membership in these groups is recorded in a treatment variable.
- Response variable compared across the two groups.
- (But really, there can be multiple "treatments" and multiple "controls".)

Issues

- How were the data collected? (Study design.)
- Who sponsored the study?
- Were results peer reviewed?
- What was the scientific context?

Why Causal Inference is Hard

- Confounding Factors
- "Confounding means a difference between the treatment and control groups -- other than the treatment--which affects the response"
  --David Freedman (Statistical Models)



Now, Genes confound the causal association.

Data collection methods

- Anecdotes
- Observational Studies
- Controlled Experiments

Word of Mouth Spreading Peanut Milk

Customers say they can vouch for the concoction's health benefits and don't need scientific proof.
By John M. Glionna
Times Staff Writer

May 17, 2006

"Listen, I'm not crazy," the Macy's worker said. "I know this stuff has made me a healthier woman."

Name that data collection method:

Anecdote

Observational

- Observational studies are those in which the subjects place themselves into treatment groups.
- Obviously, this may happen without their awareness.

Observational?

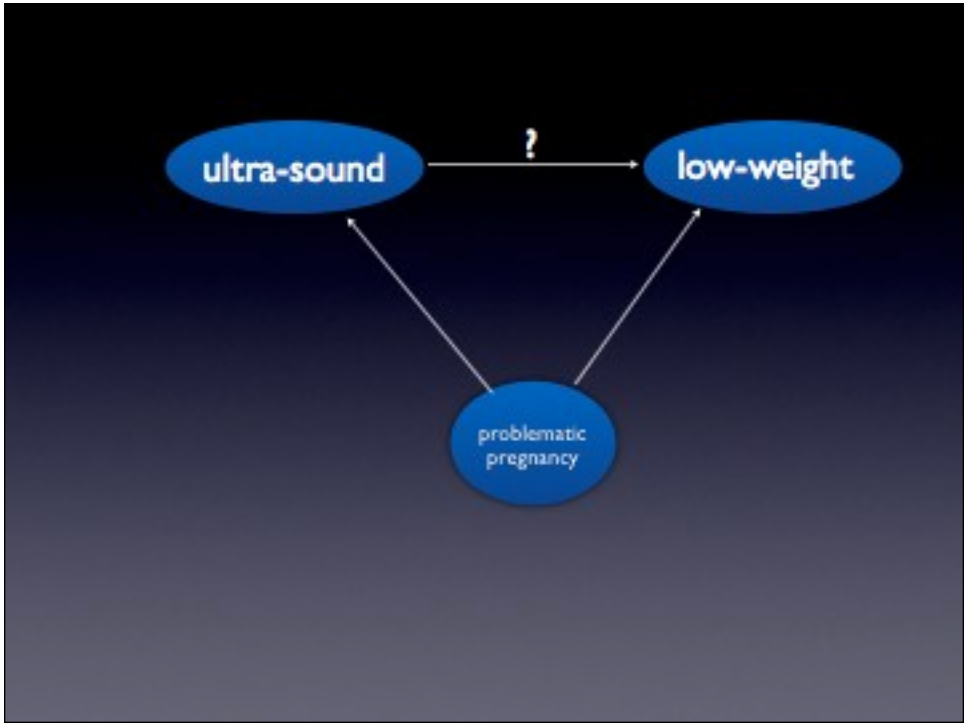
- Cellphones cause brain cancer.
- Identify treatment and response variables.
- Observational? why?
- Yes, because no researcher could control a person's cell phone exposure over time.

In an observational study, researchers found that women who had ultrasound during pregnancy were more likely to have an underweight baby than women who did not.

Conclusion: ultrasound can be harmful towards pregnancy

1. True
*** 2. False

Can you suggest a confounding factor?



Those who drink in moderate amounts are healthier, on average, than those who abstain and those who are heavy drinkers.

This is a fact.

Does moderate drinking improve health?

What kind of study and why?

Observational

Can you name some potential confounding factors?

Social Status, Prior Health, Diet

- We cannot say "Drinking in moderation lengthens life".
- We can only say "Moderate drinking is associated with long life.

Fact of Life

- One can never know if there are confounding factors in an observational study.
- Researchers can eliminate, but there could always be a nagging doubt that we just haven't thought of the right one.

The secret to successful causation

- Make sure the groups you are comparing are similar *in every way* except for the value of the treatment variable.
- Said differently: make sure there are no confounding factors.
- Easier said than done. But here's one way of doing:

Randomized, Controlled Experiments

- Random assignment assures similar groups.
- If sample sizes are large, both groups will be "the same", thus eliminating confounders.
- If sample sizes are not large, both groups are the same "on average".

Randomized

- Only in a randomized, controlled experiment can we make causal inference.
- and even then, we need to be aware of problems.

Why?

- Because only in a randomized, controlled experiment are the treatment and control groups the same.
- Thus, if the response variable is different the only explanation is the treatment.

Intent to Treat

Health Insurance Plan study, New York (circa 1960)

Can mammagram screening prevent death from breast cancer?

	Group size	Deaths	Deaths
		Number	Rate
Treatment			
Screened	20,200	23	1.1
Refused	10,800	16	1.5
Total	31,000	39	1.3
Control	31,000	63	2.0

Conclusion

- Because no confounding factors remain (that we know of), we must conclude that the reason that breast cancer rates were lower in the treatment group was that screening prevents breast cancer.

What every student should know & understand

- Data beat anecdotes.
- Random assignment in comparative experiments allows causal conclusions to be drawn.
- Random sampling allows generalization to the population

Guidelines for Assessments and Instruction in Statistics Education (GAISE), College Report
www.amstat.org/education/gaise

	Random assignment	No Random Assignment
Random Sample	causality can be extended to the population	No causality, but an association can be extended to population.
No Random Sample	Causality, but only for the sample.	no statistical inference

Students should recognize

- Common sources of bias in surveys and experiments.
- How to determine when cause-and-effect can be inferred, based on how data were collected.

Guidelines for Assessments and Instruction in Statistics Education (GAISE), College Report
www.amstat.org/education/gaise

Students should know

- How to critique news stories and journal articles that include statistical information, including identifying what's missing in the presentation and the flaws in the studies or methods used to generate the information.

Guidelines for Assessments and Instruction in Statistics Education (GAISE), College Report
www.amstat.org/education/gaise

What to Teach

- How to distinguish observational studies from controlled experiments.
- How to identify confounders and explain why they confound.
- Don't rush to conclusions based on a single study.
- Controlled, randomized experiments can go wrong.

What to Teach

- Give students headlines. Ask them whether the headline is making a claim for causation or for association.
- Students have difficulty telling these apart because our everyday language blurs the distinction.

What To Teach

- Amazing claims require amazing evidence!
- (attributed to "Amazing Randi" or possibly Carl Sagan.)

homeopathy

Los Angeles Times, 12/6/2010

- The claim: Hyland's Cold 'n Cough 4 Kids will allow children to get over colds faster or prevent colds.
- Solution so heavily diluted that there are no molecules of active ingredients in the solution.

homoepathy

- Proponents say that solution retains a "memory" of active ingredients.
- But as [Prof. Gleason, chemistry] explains, every molecule of water in our bodies has been enough other places --oceans, sewers--to make any "memories" hopeless jumbled.

homoepathy

- So even if a randomized, controlled study concluded that there was a benefit to the Cold 'n Cough, Prof. Gleason would need to see more, because if the study was true, then many things we have longed believed about chemistry are false.

Questions for students

- What is the research question?
- What is their answer?
- How were data collected?
- Are conclusions appropriate for data collection methods?
- To what population, if any, do conclusions apply?
- Have results been replicated? Or are they "amazing"?

Gould & Ryan, 2012

The Ghost Map, Steven Johnson

London Cholera Outbreak, 1854

- Cholera kills quickly
- Primary symptom is severe diarrhea
- Modern treatment involves an IV to provide hydration
- Competing Theories for transmission: air vs water (germs). Air was the predominant view.

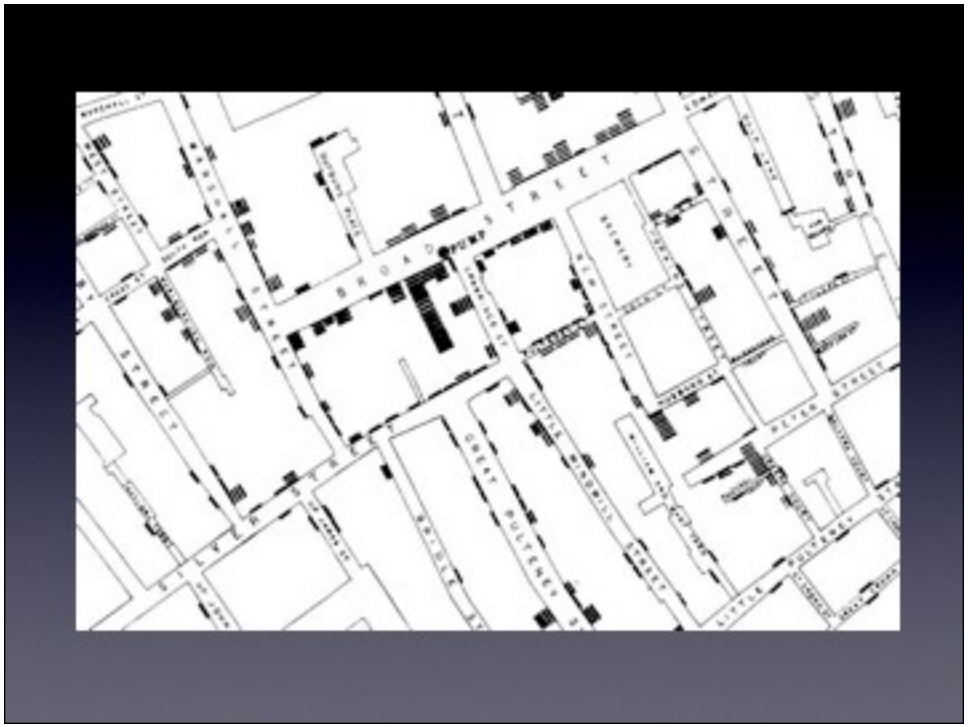
Golden Square, Sept 1854

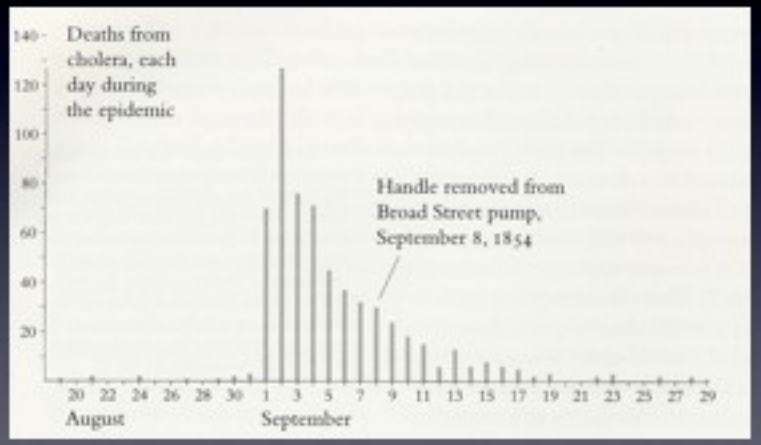
- Unusually intense, deadly outbreak
- 500 deaths in just a few days
- Local panic made neighbors flee
- John Snow, a physician, believed cholera spread through polluted water supplies

- Bleeding
- Brandy
- Saunder's Anti-Mephitic Fluid
- Castor Oil
- And from the surgeon-in-chief of the city police:

"If any corroborative testimony of [the efficacy of laudanum] be further required, I would ask those who might be skeptical of its merits to call at any one of the police stations in the city of London, where a supply of the medicine is kept and satisfy themselves of the estimation in which it is held by the members of the force"

The Ghost Map, Steven Johnson


BROAD STREET
GREAT PULTENEY STREET


Deaths from cholera, each day during the epidemic
Handle removed from Broad Street pump, September 8, 1854
140
120
100
80
60
40
20
20 22 24 26 28 30 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
August
September

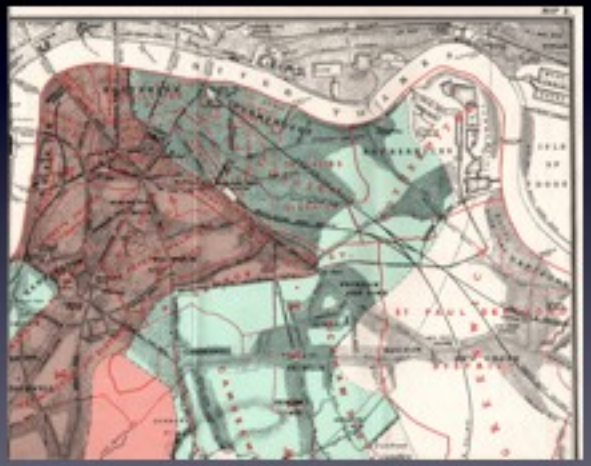


TABLE IV.

Water Companies.	Sources of Supply.	Aggregate of Districts supplied chiefly by the respective Companies.		Deaths in 100,000 inhabitants.
		Population.	Deaths by Cholera in 13 wks. ending Nov. 19.	
(1) Lambeth and (2) Southwark and Vauxhall.	Thames, at Thames Ditton and at Battersea . . .	346,363	211	61
Southwark and Vauxhall .	Thames, at Battersea . . .	118,267	111	94
(1) Southwark and Vauxhall, (2) Kent .	Thames, at Battersea; the Ravensbourne, in Kent, & ditches and wells .	17,805	19	107


JOHN SNOW

Resources

- Paul, C. (2002), Finding the Findings Behind the News, STATS: The Magazine for Students of Statistics
- Chance news, http://www.causeweb.org/wiki/chance//index.php/Main_Page
- http://jonathan.mueller.faculty.noctrl.edu/100/correlation_or_causation.htm

- www.stat.ucla.edu/~rgould

Prof. Development

- USCOTS: May 19-21, Raleigh-Durham, NC.
- Pre-USCOTS workshop for two-year college instructors of statistics
- see www.causeweb.org for details as released