Seeing Through Statistical Studies

Professor Jessica Utts
Department of Statistics
University of California, Irvine
June 21, 2012

AMERICAN STATISTICAL ASSOCIATION

UNIVERSITY OF CALIFORNIA, IRVINE

Topics

1. Surveys and polls – good and not so good
2. Correlation versus causation: how to know
3. Relationships - real or just by chance?
4. Many-fold increase in risk versus high risk
5. Why many studies find conflicting results
6. Poor intuition about probability and chance

Note: Many examples in this talk are from my textbook, *Seeing Through Statistics*

The Beauty of Surveys and Polls When Done Right

With proper sampling methods, based on a sample of about 1000 adults we can almost certainly estimate, to within 3%, the percentage of the entire population who have a certain trait or opinion.

This result does *not* depend on how large the (large) population is. It could be tens of thousands, millions, billions….

(1000 and 3% is just an example; % depends on the size of the sample. 3% = margin of error)

Estimating a Population Percent from a Sample Survey: Margin of Error

For a properly conducted sample survey:

The sample percent and the population percent rarely differ by more than the **margin of error**. They do so in fewer than 5% of surveys (about 1 in 20).

(Conservative) Margin of error $\cong \dfrac{1}{\sqrt{n}} \times 100\%$

where ***n*** is the number of people in the sample.

Example: June 7-11, Reuters Poll of ***n* = 848 registered voters asked:**

"If the election for U.S. Congress were held today, would you vote for the Democratic candidate or the Republican candidate in your district where you live?"

	Democrat	Republican	Neither/Unsure
Results:	47%	44%	9%

Margin of error is $\frac{1}{\sqrt{848}} = 0.034 \text{ or } 3.4\%$

So results for the two parties are within the margin of error.

Bias: How Surveys Can Go Wrong

Results based on a survey are **biased** if the methods used to obtain those results would consistently produce values that are either too high or too low.

Selection bias occurs if the method for selecting participants produces a sample that does not represent the population of interest.

Nonparticipation (nonresponse) bias occurs when a representative sample is chosen but a subset cannot be contacted or doesn't participate (respond).

Response bias (**biased response**) occurs when participants respond, *but* they provide incorrect information, intentionally or not.

Extreme Selection Bias: *A meaningless poll*

Responses from a **self-selected group** or **volunteer sample** usually don't represent any larger group.

Example: "Do you support the President's economic plan?"

	TV Station call-in poll	Proper survey
Yes (support plan)	42%	75%
No (don't support plan)	58%	18%
Not sure	0%	7%

Those *dissatisfied* more likely to respond to TV poll. Also, it did not give the "not sure" option.

Sources of *Response* Bias

1. *Deliberate bias*
2. Unintentional bias
3. Desire to please
4. *Asking the uninformed*
5. Unnecessary complexity
6. Ordering of questions
7. Confidentiality and anonymity

Deliberate Bias

Questions can be deliberately worded to support a certain cause.

Example: *Estimating what % think abortion should be legal*

- Anti-abortion group's question: "***Do you agree that*** abortion, the murder of innocent beings, should be outlawed?"
- Pro-choice group's question: "***Do you agree that*** there are circumstances under which abortion should be legal, to protect the rights of the mother?

Appropriate wording should not indicate a desired answer.

Asking the Uninformed

People do not like to admit they
don't know what you are talking about.

1995 Washington Post poll #1:

1000 randomly selected respondents asked this question about the *non-existent* 1975 Public Affairs Act:

"Some people say the 1975 Public Affairs Act should be repealed. Do you agree or disagree that it should be repealed?"

43% of sample expressed an opinion –
with 24% agreeing and 19% disagreeing.

Source: Morin, *Wash. Post Weekly,* 10-16 April 1995, p. 36.

Example, continued…

1995 Washington Post poll #2:

Two groups of 500 randomly selected respondents.

Group 1: "President Clinton said that the 1975 Public Affairs Act should be repealed. Do you agree or disagree?"

Group 2: "The Republicans in Congress said that the 1975 Public Affairs Act should be repealed. Do you agree or disagree?"

Group 1: 36% of Democrat respondents agreed, only **16% of Republican** respondents agreed.

Group 2: 36% of Republican respondents agreed, only **19% of Democrat** respondents agreed

Topic 2:
Can cause and effect be concluded?

Randomized experiment:
Researchers

- ***Create*** differences in groups
- *Observe* differences in response

Example:

Randomly assign women to take HRT or not, observe and compare heart disease rates

Observational study:
Researchers

- ***Observe*** differences in groups
- *Observe* differences in response

Example:

Ask women if they take HRT or not, observe and compare heart disease rates

Can cause and effect be concluded?

- Observational studies are sure to have *confounding variables* that can't be separated from the variables of interest.
- Randomized experiments help to even out confounding variables across groups.

Example:

Explanatory variable: Hormones or not

Response variable: Heart disease or not

Possible confounding variables:

Level of health care, amount of exercise, diet, general attitude toward one's health, etc.

Examples for Discussion

Explanatory variable and response variable?

Observational study or randomized experiment?

Possible confounding variables?

- *A NEJM study found that members of households with guns were 2.7 times more likely to experience a homicide than those in households without guns.*
- *The Abecedarian Project (UNC) randomly assigned poor infants to receive full-time, educational child care until kindergarten, or not. Those with child care were almost 4 times as likely to graduate from college (23% vs 6%); there were many other differences too.*

Assessing possible causation

Some features that make causation plausible:

- There is a reasonable explanation for how the cause and effect would work.
- The association is consistent across a variety of studies, with varying conditions.
- Potential confounding variables are measured and ruled out as explanations.
- There is a “dose-response” relationship.

Topic 3: Real Relationships or Chance? Does eating cereal produce boys?

- Headline in *New Scientist*: "Breakfast cereal boosts chances of conceiving boys" Numerous other media stories of this study.
- Study in *Proc. of Royal Soc. B* showed of women who ate cereal, 59% had boys, or women who didn't, 43% had boys.
- Problem #1: Headline implies eating cereal *causes* change in probability, but this was an observational study. Confounding likely!

The Problem of Multiple Testing

- The study investigated 132 foods the women ate, at 2 time periods for each food = 264 possible tests!
- By chance alone, *some* food would show a difference in birth rates for boys and girls.
- Main issue: Selective reporting of results when many relationships are examined, not adjusted for multiple testing. Quite likely that there are "false positive" results.

Common Multiple Testing Situations

- *Genomics*: "Needle in haystack" – looking for genes related to specific disease, testing many thousands of possibilities.
- *Diet and cancer*: Ask cancer patients and controls about many different dietary habits.
- *Interventions* (*e.g. Abecedarian Project**): Look at many different outcomes and compare them for the groups that had different interventions.

  *In this case there were too many differences to be explained by chance.

Multiple Testing: What to do?

If you read about a study and suspect multiple testing is a problem:

- There are statistical methods for handling multiple testing. See if the research report mentions that they were used.
- See if you can figure out how many different relationships were examined.
- If many significant findings are reported (relative to those studied), it's less likely that the significant findings are false positives.

Topic 4:
Avoiding Risk May Put You in Danger

- In 1995, UK Committee on Safety of Medicines issued warning that new oral contraceptive pills "increased the risk of potentially life-threatening blood clots in the legs or lungs by twofold – that is, by 100%" over the old pills
- Letters to 190,000 medical practitioners; emergency announcement to the media
- Many women stopped taking pills.

Clearly there is increased risk, so what's the problem with women stopping pills?

Probable consequences:

- Increase of 13,000 abortions the following year
- Similar increase in births, especially large for teens
- Additional $70 million cost to National Health Service for abortions alone
- Additional deaths and complications probably far exceeded pill risk.

Risk, Relative Risk and Increased Risk

- The "twofold" risk of blood clots was a change from about 1 in 7000 to 2 in 7000 – not a big change in *absolute* risk, and still a small risk.
- *Absolute risk, relative* risk and *increased* risk
  - *Absolute* risk: The actual risk; in this case 2 out of 7000 were likely to have a blood clot
  - *Relative* risk: How much the risk is *multiplied* when comparing two scenarios, double in this case
  - *Increased* number at risk: Change in number at risk; from 1 in 7000 to 2 in 7000 in this case

Considerations about Risk

- Changing a behavior based on relative risk may *increase* overall risk of a problem. Think about trade-offs.
- Find out what the absolute risk is, and consider relative risk in terms of additional *number* at risk

Example: Suppose a behavior doubles risk of cancer

Brain tumor: About 7 in 100,000 new cases per year, so adds about 7 cases per 100,000 per year.

Lung cancer: About 75 in 100,000 new cases per year, so adds 75 per 100,000, more than 10 times as many!

Topic 5:
Why Do Studies Find Conflicting Results?

Ioannidis (2005) looked at replication:

- 45 high-impact medical studies in which treatments were found to be effective
  - Each published in top medical journal, and had been cited more than 1000 times
  - Studies were repeated with same or larger size, and same or better controls for 34 of them.
- How many do you think replicated original result of effective treatment? All? Most?

Conflicting results, continued

The 45 studies included 6 observational studies and 39 randomized controlled trials.

Replication results:

- Only 20 of the 45 attempted replications were successful (i.e. found the same or better effect)
- Of the 6 observational studies, 5 found smaller or reversed effects (83%).
- Of the 39 randomized experiments, 9 found smaller or reversed effects (23%).

Possible explanations

Ioannidis suggests these explanations:

- Confounding variables in observational studies
- Multiple testing problems in the original studies
- Multiple researchers looking for a positive finding; by chance alone, someone will find one

Other possible explanations:

- Different conditions or participants (different ages, incomes, etc.) in the two studies
- Successful replications less likely to be published than unsuccessful ones – "nothing new"

Topic 6:
Poor intuition about probability and chance

- William James was first to suggest that we have an intuitive mind and an analytical mind, and that they process information differently.
- Example: People feel safer driving than flying, when probability suggests otherwise.
- Psychologists have studied many ways in which we have poor intuition about probability assessments.

Example: Confusion of the Inverse

Gigerenzer gave 160 gynecologists this scenario:

- About 1% of the women who come to you for mammograms have breast cancer (bc)
- If a woman has bc, 90% chance of positive test
- If she does not have bc, 9% chance of positive test (false positive)

A woman tests positive. What should you tell her about the chances that she has breast cancer?

Answer choices: Which is best?

- The probability that she has breast cancer is about 81%.
- Out of 10 women with a positive mammogram, about 9 have breast cancer.
- Out of 10 women with a positive mammogram, about 1 has breast cancer.
- The probability that she has breast cancer is about 1%.

Answer choices and % who chose them

- 13% chose "The probability that she has breast cancer is about 81%."
- 47% chose "Out of 10 women with a positive mammogram, about 9 have breast cancer." [Note that this is 90%.]
- 21% chose "Out of 10 women with a positive mammogram, about 1 has breast cancer." [Note that this is 10%.]
- 19% chose "The probability that she has breast cancer is about 1%."

Correct answer is just under 10%!

Let's look at a hypothetical 100,000 women:

	Test positive	Test negative	Total
Disease	**900 (90%)**	100	1000 (1%)
No disease	8910 (9%)	90,090	99,000
Total	**9810**	90,190	100,000

Physicians confused two probabilities:

- Probability of positive test, *given* cancer (90%)
- Probability of cancer, *given* positive test = 900/9810 = 9.2%

Confusion of the inverse: Other examples

Cell phones and driving (2001 study):

- Given that someone was in an accident:
  - Probability that they were using cell phone was .015 (1.5%)
  - Probability that they were distracted by another occupant was .109 (10.9%)
  - Does this mean other occupants should be banned while driving??
- What we really want is probability of being in an accident, *given* that someone is on a cell phone, much harder to find!

Confusion of the inverse: DNA Example

Dan is accused of crime because his DNA matches DNA at a crime scene (found through database of DNA). Only 1 in a million people have this specific DNA. Is Dan surely guilty??

Suppose there are 6 million people in the local area population, so about 6 have this DNA. Then:

- Probability of a DNA match, *given* that a person is innocent is only 5 out of 6 million – very low!
- But... probability that a person is innocent, *given* that his DNA matches is 5 out of 6 – very high!

Other Probability Distortions

- Coincidences have higher probability than people think, because there are so many of us and so many ways they can occur.
- Low risk, scary events in the news are perceived to have higher probability than they have (readily brought to mind).
- High risk events where we have think we have control are perceived to have lower probability than they have.
- People place less credence on data that conflict with their beliefs than on data that support them.

Some Useful Websites

http://coalition4evidence.org

Nonprofit, nonpartisan organization to increase government effectiveness through rigorous evidence of what works in areas of social policy

http://www.harding-center.com

Harding Center for Risk Literacy (*Gerd Gigerenzer*)

http://stats.org (George Mason University)

Nonprofit, nonpartisan, discussions of use and abuse of statistics in the media and public policy

QUESTIONS?

Contact info:

jutts@uci.edu

http://www.ics.uci.edu/~jutts

AMERICAN STATISTICAL ASSOCIATION

UCIrvine UNIVERSITY OF CALIFORNIA, IRVINE