Spinning Heads and Spinning News

Media's Use of Statistics

Rebecca Goldin,
STATS and Department of Mathematics
George Mason University

Numbers describe our world. Statistics shape our politics and policies. But can they always be trusted? STATS mission is to examine how numbers are distorted and statistics are misunderstood in the media and in society.

"Interesting, unexpected analyses" - **The Washington Post**

Statistical Assessment Service (STATS)

Director of research	Rebecca Goldin, Ph.D.
Editor	Trevor Butterworth

www.stats.org

Taking a critical look

Economics/Money
Crime
Social behavior
Budget
Sports
Polls/Opinion
Education
Public Health/Health care

Common Misunderstandings

The meaning of *statistically significant*
Causation versus correlation
Relative risk versus absolute risk
Confounding factors
Scales and orders of magnitude
Margin of error
Consensus of scientists versus individual research results

Causation or Correlation

The New York Times reported, "Lengthy television viewing in adolescence may raise the risk for depression in young adulthood, according to a new report." Also picked up by *US News and World Report*

The article noted "while about 6 percent of those who watched under three hours a day were depressed, more than 17 percent of those who watched more than nine hours a day had depressive symptoms."

Who is watching TV for nine hours a day?

Other Recent Television Research

Ads increase the viewing pleasure of television

But television is also linked to childhood obesity (the link is attributed to the ads, according to several news sources).

Recently, television was causally linked to early initiation of sexual activity in girls (*MSNBC, WebMD,USA Today, Assoc. Press*). Only *Washington Post* noted it might not be causal.

Often, there are obvious reasons that the correlation might be observed, without television being the cause.

The Obesity Epidemic

• Television
• Sugar soda (refined flour)
• Meat consumption
• Poverty
• Reduced physical activity
• Public transportation
• Time Spent Eating Per Day

Time Spent Eating Per Day v. National Obesity Rate

Washington Post:
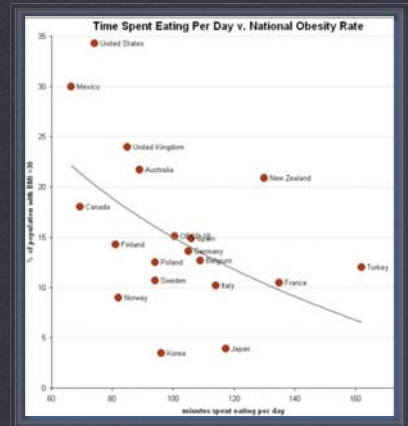
"Those who described their health as "excellent" -- people who presumably had relatively little experience pursuing medical care or submitting claims -- were almost twice as likely as those in good, fair or poor health to rate their private health insurance as excellent."

Percentage of people in excellent health

Happy with their insurance	Unhappy with their insurance
63%	32%

SAMPLE

But the chart indicated the results of a survey asking those people who are happy with their insurance how good their health is.

"Health Insurance Industry Spins Data in Fight Against Public Plan"

The Magic Coin

Heads: Republican, Tails: Democrat

Worked for previous 15 out of 17 elections. Chance of just above 0.1% to do so well by chance.

How I found it…..

In a room of 1000 flippers, there's a 70% chance of finding a magic coin.

The problem is that simple coincidences are everywhere, and our minds are constantly looking for patterns. After we've found a magic coin, we don't always wonder whether it was one of thousand regular coins.

What's wrong with the anecdote?

A kid commits suicide while taking anti-depressants. Is it just a chance occurrence or something worse?

Coincidences are everywhere. How do we know which are true associations/correlations?

Hierarchy of Study Design

Controlled (also called experimental);

A controlled study on a drug would take two groups of people (let's call them A and B) and give set A the drug and give set B a placebo. None of the participants would know which group they were in, so no one would know if he or she got the "real" drug. The researchers would then record whether set A really has benefits compared to group B. The study is called controlled because each element of it is controlled by the people designing the study.

Observational (also called epidemiological); and

An observational study would search out data about people already using the drug, and compare it to data on comparable people who are not using the drug. It is called observational because the tactic is to observe the effect of the drug (or an activity or a lifestyle) in the general population, without recruiting participants into a controlled environment.

Case-control.

Case-control studies are typically used to find side effects rather than to evaluate the effectiveness of a drug. They test the relationship between two occurrences to see if they are correlated by comparing/contrasting people with a certain trait, against similar people who do not exhibit the same trait.

Confound it!

Primetime Live Poll: More Republicans Satisfied With Sex Lives Than Democrats

* Very satisfied with their sex lives: 56 percent of Republicans compared to 47 percent of Democrats

* Wore something to enhance their sex lives: 72 percent of Republicans compared to 62 percent of Democrats

* Faked an orgasm: 28 percent of Republicans compared to 33 percent of Democrats

Didn't control for the confounder of gender. More women are Democrats.

TABLE 3. Percentage of Ever Breastfed and ORs for Postneonatal Deaths

| | Original Samples | | | | | SUDAAN-Adjusted | | | | |
| | Live births | | Postneonatal Deaths | | Adjusted* OR (95% CI) | Live Births | | Postneonatal Deaths | | Adjusted* OR (95% CI) |
	n	% Ever Breastfed	n	% Ever Breastfed	Ever/Never Breastfed	N	% Ever Breastfed	N	% Ever Breastfed	Ever/never Breastfed
Total	7740	39.7	1204	31.2	0.76 (0.65–0.88)	3 186 497	53.4	9145	37.7	0.79 (0.67–0.93)
Race specific										
Black	4038	26.0	561	16.0	0.67 (0.52–0.87)	549 360	27.9	2498	16.8	0.69 (0.53–0.90)
Nonblack	3702	54.7	643	44.3	0.79 (0.65–0.96)	2 637 137	58.7	6647	45.6	0.81 (0.66–0.98)
Birth weight specific										
<2500 g	2015	27.7	287	24.7	0.87 (0.60–1.37)	200 443	36.1	2002	31.4	0.97 (0.64–1.47)
≥2500 g	5725	43.9	917	33.2	0.73 (0.61–0.86)	2 986 054	54.5	7143	39.5	0.76 (0.64–0.91)
Death cause specific										
Infections	-	-	255	30.2	0.75 (0.55–1.02)	-	-	1914	37.0	0.76 (0.54–1.07)
Injuries	-	-	126	28.6	0.67 (0.43–1.05)	-	-	971	31.9	0.59 (0.38–0.94)
SIDS	-	-	591	31.6	0.77 (0.63–0.95)	-	-	4514	38.3	0.84 (0.67–1.05)
Others	-	-	232	32.3	0.76 (0.54–1.07)	-	-	1745	40.4	0.81 (0.56–1.16)

* Adjusted for mother's age, education, and smoking during pregnancy and infant's gender, race (except for race subgroup analyses), birth weight (except for birth weight subgroup analyses), congenital malformation reported at birth, live birth order, plurality, and WIC status.

How to read a table
Confidence Intervals, Confounders, Odds Ratio

Writing about risk

AP Probe Finds Drugs in Drinking Water

- A vast array of pharmaceuticals — including antibiotics, anti-convulsants, mood stabilizers and sex hormones — have been found in the drinking water supplies of at least 41 million Americans, an Associated Press investigation shows.

- Officials in Philadelphia said testing there discovered 56 pharmaceuticals or byproducts in treated drinking water, including medicines for pain, infection, high cholesterol, asthma, epilepsy, mental illness and heart problems. Sixty-three pharmaceuticals or byproducts were found in the city's watersheds.

- Anti-epileptic and anti-anxiety medications were detected in a portion of the treated drinking water for 18.5 million people in Southern California.

- Researchers at the U.S. Geological Survey analyzed a Passaic Valley Water Commission drinking water treatment plant, which serves 850,000 people in Northern New Jersey, and found a metabolized angina medicine and the mood-stabilizing carbamazepine in drinking water.

- Three medications, including an antibiotic, were found in drinking water supplied to Tucson, Ariz.

- A sex hormone was detected in San Francisco's drinking water.

- The drinking water for Washington, D.C., and surrounding areas tested positive for six pharmaceuticals.

What wasn't said

GET WORRIED! THIS IS TERRIBLE! DRUGS ARE BAD!

- **No discussion of orders of magnitude.** Drugs were found in concentrations ranging from parts per billion to parts per trillion. That's like saying something costs somewhere between $1,000 and $1,000,000.

- **No discussion of effects in minute quantities.** The dose makes the poison. If it weren't, naturally-occurring chemicals in vegetables would kill us.

- **No comparisons.** How many parts per billion of oil or gasoline or other kinds of contaminants are normally found in drinking water? (some of this water comes from treated sewage; scary to think about what lurks in parts per trillion levels there!)

RU-486. When the moral and the statistical collide.

- What are the choices for a pregnant woman and the risks involved?

- Are you reporting from a public health perspective, an individual's perspective, or from your moral gut?

- Notice in the article that the risks are not mentioned, either for using or not using the abortion pill.

Political actions based on "science"

- Move to ban phthalates from toys – when there is little evidence that the route of exposure is through toys.

- Recommendations regarding alcohol consumption for pregnant and nursing women

- Educational programs designed based on poor data (No Child Left Behind)

- Abstinence Education

Averages are tricky

Mean (also called the average or the norm);

The mean is the classic average. It is the sum of all the numbers, divided by the number of numbers you sum. For example, the average of 5, 6, and 8 is 19/3, or 6 1/3. Averages work well when there aren't too many extreme values. For example, the average height of a 9-year-old boy in the United States is about 46 inches.

Median (sometimes called the average);

The media of a set of numbers is a value such that half of the numbers are larger, and half of the numbers are smaller. For example, the median of 3, 5, 6, 7, 19, 87, 92 is 7. Notice that the average would be much higher. The median is a good number to use when the data include some very large or very small outliers. In this case, the median describes the "typical" person better than the mean. For example, the median income in the United States is about $25,000 per year for a family of four.
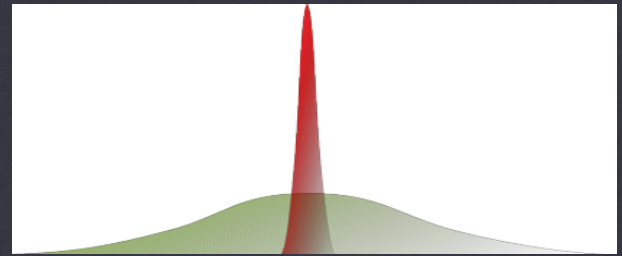
Mode

The mode is the most common number in a data set. For example, the mode for the number of feet in the United States is 2. The average is slightly lower.

The Normal (or Bell) Curve and Standard Deviation
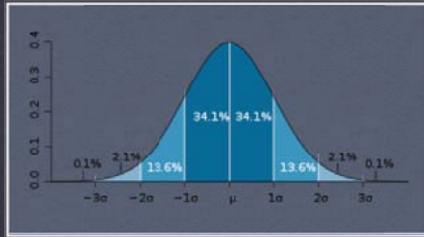
Two normal curves
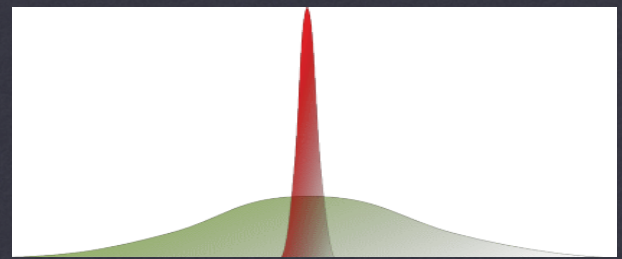

Standard deviation

The standard deviation is a measurement for a set of data in a normal distribution. It tells you how much the data *vary*. One standard deviation from the mean captures about 68% of the data and two standard deviations captures about 95% of the data. For example, about 95% of kids have IQs within two standard deviations of the norm.

When do differences in standard deviation arise

- Race-based medicine

- Educational -- comparing schools, comparing educational techniques, comparing boys to girls

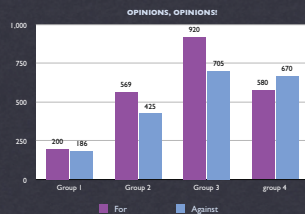- Economics - investment risk

- Health care reform

Differences in SD are socially signficant

Similar averages, different standard deviations: white boys versus girls on NCLB math tests

Polling Problems

- **Undecided**
- **Likely voters?**
- **Representative sample**
- **Margin of sampling error**
- **Bias in the question**
- **Truthfulness**

A poll gone foul, and the meaning of margin of error

A survey by the Physicians Foundation found that 49 percent of doctors planned to cut back or close their practices. A statistician endorsed the claim that the margin of error was one percent.

USA Today headlined their article, "Primary care doctors in short supply."

Turns out the response rate was 4 percent.

Margin of Error: what it is and what it's not

- The margin of error tells us the range of values that we can be "confident" in. We can be **95 percent confident** that the whole population is within the margin of error of what we measured. For example, if a poll finds that 53 percent of women claim their insurance does not cover brand name birth control pills with a margin of error of 3 percentage points, then we can be 95 percent confident that somewhere between 50 and 56 percent of all women's health insurance does not cover brand name birth control.

- Under some assumptions, it depends only on the number of people polled or sampled. A margin of error of about 3 percent is obtained when just over 1,050 people are polled.

- Margin of error only measures error due to "randomly" choosing non-representative respondents. It cannot measure response bias (as in the previous example), bias in the question, bias in who/how you poll, etc.

Statistical Topics

Increasing Numeracy in the Media

Sensitivity versus Specificity

Lie Detectors (Washington Post)

10,000 people take a lie detector test, of which
10 are spies (and the rest honest folk).
Research shows that just under 1,600 people would fail the test, including 8 of the spies. The other two spies would pass.
Suppose that exactly 1,592 out of 9,990 innocent people fail the test.
These are false positives – people who tested positive for lying when they were not.
The false positive rate is 1592/9990 or about 16 percent.
                    Sensitivity is 84 percent.
On the other hand, eight out of ten liars failed the test.
                    Specificity of 80 percent.
With 10,000 people taking the test, of which 10 are spies, the percentage of spies in the failed pool is 8 out of 1600, or .5 percent. That makes it pretty hard to find the spies!

When does this come up in medical reporting?

Medical testing: its worth quite generally (mammograms, prostate tests)
Decisions regarding the reaction to a result from a test, such as ultrasound abnormalities in fetuses during pregnancy

Scientific Culture

- News agencies hear of a surprising result and think that it makes for a great story

- Scientists hear of a surprising result and wonder, "What went wrong?"

- There are some "activist scientists" who promote their work shamelessly.

- Survey of the Society of Toxicology (a reflection of scientific culture and perception).

What is the media up against?

STATS president and George Mason University Professor Bob Lichter: survey of members of the Society of Toxicologists about chem risk

WebMD and Wikipedia are seen as more reliable than traditional news.

90% say media coverage of risk lacks balance
97% say media doesn't distinguish good studies from bad studies
96% say the media doesn't distinguish correlation from causation.
74% say media pays too much attention to individual studies; 73% say to individual scientists.
80% say leading media overstate chemical risks.
15% and under say leading media accurately portrays chemical risk

What does STATS *do*?

* Write op-eds
* Speak on radio shows
* Letters to the editor
* Interview for television programs on topics of interest
* Practice in-depth, data driven journalism
* Advise journalists
* Run workshops
* Provide educational materials (FAQ)
* Run website with material plus blog
* Testify to Congress
* Poll the experts and publish the results

Good medical reporting

Establish scientific consensus. Consider what government agencies say and scientists say with more weight than environmental groups and industry. Ask other researchers if they agree with the methods and results. Look for results of panelists of experts. Search the literature for reviews that talk about the literature as a whole.

Evaluate the QUALITY of scientific research. How large are the studies? In the case of toxic substances, are the damning studies examining the potential hazard with consideration to exposure route? In the case of "lifestyle choices" (alcohol, food, exercise, etc.), did the researchers control for obvious confounding factors? For individual studies, were the results statistically significant? What was the study design and what might it have left out? Have mechanisms been proposed?

Regard press releases suspiciously. Scientists are also trying to promote their work.

Report sources. Make PubMed references/ID numbers. Put links and report where a study is published. Link and mention useful sites for more information.

Good medical reporting

Get familiar with study methods for each discipline. The methods of research in toxicology are different than in biomedical work. Learn the principles, such as whether method of dosing matters (oral, subcutaneous, intravenous, etc), what the most important confounders are and why, etc.

Understand sources. Industry funded research often follows good laboratory practice, and should not be dismissed without investigation. Importantly, a lot of industry research is by law the sources of proving pharmaceutical drugs are safe. Government laboratories are the most revenue neutral, but they have restrictions on their scope as well. Certain industries, foundations, and environmental groups have specific aims that are at conflict with science.

Be careful not to count all studies as equal. The argument that "17 studies show one thing and only 2 show another" is fallacious, because the two could be much better studies.

Get disinterested expert opinions.

Read the abstract, results, and conclusions section of a research article. Many interesting stories develop this way!

2008 Dubious Data Awards

Baby Bottle Baloney. *USA Today* and *Milwaukee Sentinel Journal* claim chemicals leeching from baby bottles cause obesity. The lead scientist's work has been dismissed by independent risk analyses in Europe.

Acupuncture and Fertility. *Fox News, MSNBC,* the *Examiner,* and the *Washington Times* reported that acupuncture is a useful remedy for fertility. They confused odds ratio with relative risk. The benefits were much smaller than reported.

Rainfall and Autism: The Sound of Statistical Noise. *USA Today* covered a study by an economist claiming that counties with high levels of rainfall in the Pacific Northwest had higher rates of autism among children. There was no direct link between rain and autism in the study.

Treats, Shoots and Misses. *New York Times* reported on a study suggesting US Health Care is dismal compared to other country's. The study was based on patient satisfaction rather than outcome.

Toxic Shower Curtain Smell. *Los Angeles Times* reported that vinyl shower curtains were a risk to our health because they contain phthalates, a family of chemicals used in plastics. Laboratory tests for the scientific report on which LA Times based its story found that the phthalates coming off the curtains were below detection level.

Latent Sexism. *The Boston Globe* reported on the latest research about the lack of women in science and engineering, and decided that it means women were embracing "the freedom to say 'no.'" As in saying "no" to careers in science. Turns out that biology wasn't counted, though 60 percent of degrees go to women.

"Dublin Inhabited by Little People." *The New York Times* reported that "A million people live inside Dublin's official borders, which are such that you can literally walk anywhere in about half an hour." This implies that the city modeled by a circle has a diameter of approximately 2 miles, making it about five times more densely inhabited than Manhattan.

To Life!

Thank you!