Common Statistical Fallacies

William M. Briggs, Ph.D.

Introduction

The public is usually informed of scientific studies by the press, and the press usually discovers its stories by scanning press releases about and journal abstracts of peer-reviewed papers. Researchers often write these papers using questionable statistical methods. Enthusiasm often trumps caution, especially in "hot" fields.

Collected here are exemplars of several common statistical fallacies, errors in reasoning, or overreaches in interpretation that caught the attention of the public, politicians and other decision-makers, and of course of scientists. The fallacies are given their street rather than Latin names so they may be identified more readily. Readers will be able to recognize in these examples matching instances from their own experience.

1. The Publish or Perish Fallacy

The July 20, 2011, Fox News *Grapevine* segment "Stars, Stripes and Republican Votes?" reported: "A new study claims a single exposure to the American flag—even among Democratic participants—shifts support toward Republican beliefs, attitudes and voting behavior." The source for this was a research paper, whose abstract read: "We report that a brief exposure to the American flag led to a shift toward Republican beliefs, attitudes, and voting behavior among both Republican and Democratic participants, despite their overwhelming belief that exposure to the flag would not influence their behavior."[1] It is a matter of great curiosity to academics why some people are not Democrats; hence this study, a sample from the healthy genre of sociological studies of political affiliation.

The authors recruited 396 people to participate in four online survey sessions, starting from before the 2008 presidential election and lasting until 8 months after. One hundred ninety-seven made it through Session 2; just 71 made it to the end. Eight more were excluded for various reasons, leaving only 63 from which to draw conclusions.

Some of the 63 received in their first surveys a small American flag printed in the corner, and some did not. All particiants were pre-selected to have polarized political views in the sense that only those "who planned to vote...where polling indicated...a significant margin separated Obama and McCain" were included in the study. Each participant was thus likely to be an ardent and not lukewarm supporter of one of the candidates.

The authors created "composite measures" of final voting intentions from various answers on the surveys. Then they regressed "centered" versions of these composite intentions on early composite intentions and used the residuals from this unduly complex model as the main measure of political attitude. They reasoned that the output from their multi-layered manipulations was superior to participants' actual answers.

Their conclusion was that a function of the flag-seers' model residuals tended to move in the direction (so the authors claim) of disapproval of Obama's 8-month-old presidency slightly more than did the residuals of the non-flag-seers. It was the flags that caused the differences, they said, and not the ridiculous statistics and purposely biased sample that accounted for the results.

The Publish or Perish Fallacy is: When you need a paper, and academics always need papers, do a survey, call the questionnaire an "instrument," then apply an unnecessarily complicated, opaque, *sui generis*, unreplicable statistical procedure to a trivially small, biased sample, but be sure the "findings" accord with received wisdom.

2. The Overconfident Academic Expert Fallacy

US News & World Report said, "Democratic political candidates can skip this weekend's July 4th parades. A new Harvard University study finds that July 4th parades energize only Republicans, turn kids into Republicans, and help to boost the GOP turnout of adults on Election Day."[2]

The abstract of the study the magazine consulted read: "Survey evidence [says] Republicans consider themselves more patriotic than Democrats,... a political congruence between the patriotism promoted on Fourth of July and the values associated with the Republican party. Fourth of July celebrations in Republican dominated counties may thus be more politically biased events that socialize children into Republicans."[3]

From the paper: "Our method uses daily precipitation data from 1920-1990 to proxy for exogenous variation in participation on Fourth of July as a child. The estimates imply that days without rain on Fourth of July in childhood increase the likelihood of identifying with the Republicans as an adult, voting for the Republican but not the Democratic candidate, and voter turnout."

Their result: "[O]ne Fourth of July without rain before age 18 increases the likelihood of identifying as a Republican at age 40 by 2 percent, the share of people voting for the Republican candidate at age 40 by 4 percent, and the share of people turning out to vote at age 40 by 0.9 percent." If this model is correct, it means a person attending all 17 parades before his 18th birthday has a 34 percent greater chance (the 34 percent is added to some mysterious baseline) of identifying as a Republican, a 68 percent greater chance of voting for a Republican, and a 15.3 greater chance of turning up to vote.

The reader will note that actual parade attendance was never measured. Rainfall was. If it rained, even only a little, in the town that was listed as a participant's hometown, the authors assumed that the particiant did not go to a parade in that hometown or anywhere else. And if it didn't rain in that hometown, the authors assumed that each hometown had a parade, that the participant was in residence at the time, and that he necessarily went to the parade.

San Francisco almost never sees rain on the Fourth of July. Therefore, if the authors' model that sunny days force people to attend patriotic Republican-oriented parades is correct, that city

should be teeming with Republicans. Is it?

The Overconfident Academic Expert Fallacy is: When an author says, "I can't think of another explanation for the observed correlation; therefore, there isn't one, and whatever I say is the cause is the cause, especially if that cause sounds newsworthy."

3. The Friend of a Friend Fallacy

The California Air Resources Board (CARB) issued a widely disseminated press release, which stated: "Three new studies released today by [CARB] reveal that exposure to airborne fine-particulate matter significantly elevates the risk for premature deaths from heart disease in older adults and elevates incidence of strokes among post-menopausal women. Heart disease is the number one killer in California and is responsible for approximately 35% of annual deaths."[4]

This press release was based in part on work contracted by CARB (contract No. 06-332) to a group led by principal investigator Michael Jerrett. The abstract of the report stated: "All-cause mortality is significantly associated with PM2.5 exposure, but the results are sensitive to statistical model specification and to the exposure model used to generate the estimates. When we applied control for residence in the largest urban conurbations, and we employed the land use regression (LUR) model, we found significantly elevated effects on all cause mortality."[5] PM2.5 represents particles less than 2.5 micrometers in diameter, i.e. dust.

The method is complicated, but basically the authors tested nine different statistical models tying several diseases to dust; some of these were coupled disease-land-use models. Land-use models attempt to guess what segments of residential land are used for. Only one model of the nine showed a "statistically significant" relationship between dust and one of the diseases. The authors concentrated all their comment on this model.

Yet the authors never, not even once, measured the PM2.5 inhalation of any person. Instead, they measured how far from highways residents might have, at one time, lived, and then used that distance to estimate "exposure," using land-use models. There was no way to check the veracity of these estimates. Strangely, suburban residents, i.e. those living far from highways, had much higher risks of heart disease deaths than did urban residents. Yet the authors stated, "The results from this investigation indicate consistent and robust effects of PM2.5."

The Friend of a Friend Fallacy is also known as the Ecological or Epidemiologist Fallacy because it is so heavily used by epidemiologists. If you can't measure what you think is the real cause, instead measure things that conceivably might be somewhat related under some imaginable circumstance and call them the cause. Nobody will remember the substitution.

4. The Everyone Else Said It Was True Fallacy

"Radon is one of the most serious environmental health risks that we face," said Univeristy of Minnesota professor Bill Angell. He explains that the colorless, odorless radioactive gas forms naturally in the ground, but when it enters your home, it is a serious problem.

"The risk of dying of lung cancer because of radon in your home is one out of 50," said Angell, "So it's an incredibly big risk."[6]

Angell's comments were based on published studies such as a Danish cohort study by Bräuner et al., whose abstract read: "We find a positive association between radon and lung cancer risk consistent with previous studies…. [T]he results of the present prospective cohort study are fully compatible with an association between residential radon and risk for lung cancer as detected in three previous meta analyses and provide important evidence at the low end of the low end of the residential dose curve."[7]

In that study, the authors measured actual exposure and outcomes of about 57,000 Danes and found the "adjusted [risk] for lung cancer was 1.04 (95% CI: 0.69–1.56) in association with a 100 Bq/m³ higher radon concentration and 1.67 (95% CI: 0.69–4.04) among non-smokers." Since the confidence intervals include 1, the classical interpretation is that radon is therefore not significantly associated with lung cancer. In fact, the authors said as much: "The role of chance cannot be excluded as these associations were not statistically significant."

The finding of no effect was contrary to expections, so the authors said, "In the present study, a number of risk factors for lung cancer were less prevalent among participants living at the higher radon concentrations, including...low fruit intake, risk occupation and traffic-related air pollution. This would result in an underestimation of the association between radon and lung cancer risk in our study."

These words were necessary to suggest that radon might still cause lung cancer even in the face of great evidence it did not. The authors felt that something had to explain the non-effect, because they were unwilling to conceive that radon (at the stated levels) might be harmless to lungs. So in their explanation they discarded the massive evidence they collected and surmised that radon was just as deadly as commonly thought.

The Everyone Else Said it Was True Fallacy is: Even though your results are the exact opposite of your belief, explain them away, then state your belief.

5. The Statistics Aren't What You Think They Are Fallacy

Here are two headlines from *The Daily Mail*, the popular English newspaper. "Bad news for chocoholics: Dark chocolate isn't so healthy for you after all," from a Jan 24, 2012, article explaining that chocolate doesn't do much for the heart. Then just three months later, on Apr 24, another headline claimed: "Eating dark chocolate is good for your heart." Both headlines drew on different peer-reviewed medical studies that concluded, using *p*-values as evidentiary markers, that chocolate was and wasn't good for hearts.

Two more headlines from this newspaper read: "Ignore all that hype about antioxidant supplements: Why daily vitamin pills can INCREASE your risk of disease" (May 21, 2012), and "The vitamin pills that actually work! How some supplements can work wonders for certain ailments" (May 27). Some of the ailments were the same in both stories. These were also based on peer-reviewed studies, using *p*-values to "prove" their contentions.

On Apr 11, 2011, a headline announced: "Women who drink four cups of coffee a day face higher risk of incontinence." Then from *Thomson-Reuters*, (the *Daily Mail* did not cover the follow-up study) a year later, on Apr 27, 2012, readers were told: "Caffeine not tied to worsening urinary incontinence." The underlying story was the same.

On Jul 29, 2004, a headline on OBGYN.net read: "Pomegranates shown to be effective for menopausal symptoms." It took eight years for the *Daily Mail* to report on Jan 24, 2012, that: "Pomegranate seed

oil 'no better than a placebo' at easing hot flashes," (a menopausal symptom). Both reports were based on peer-reviewed studies that used *p*-values as evidence.

The Statistics Aren't What You Think They Are Fallacy is also known as the *P*-values Aren't Proof Fallacy. Researchers want to know the probability that some theory is true given the evidence they have collected. This theory is then often used in developing medical practice guidelines, particularly when the theory fits expectations.

But *p*-values, the measures upon which most studies rely, and which everybody, even those who know better, take as proof of a theory when the *p*-values are less than the magic value of 0.05, do not give evidence that any theory is true.[8] Indeed, the actual definition of a *p*-value is so complicated nobody ever remembers it; all that is recalled is that *p*-values should be small.

6. The Bandwagon Fallacy

Here are quotations from various sources on the dangers awaiting us once global warming strikes. These can be multiplied indefinitely.

"[Global warming's] indirect social and political impact in poor countries may be even more far-reaching, including upheavals and civil wars—and even more witches hacked to death with machetes" (Edward Miguel, Professor of Economics, Berkeley).

"Oyster Herpes: Latest Symptom of Global Warming?" (*National Geographic*).

"Global warming 'helps coral reefs grow'" (*New Scientist*).

"Global Warming Has Devastating Effect on Coral Reefs, Study Shows" (*National Geographic*).

"We don't usually think of the Taliban and global warming in the same sentence" (Charlie Gibson, *ABC News*).

"[Global warming will cause] plankton called a coccolithophore to bloom in huge numbers" (Sharon Smith, professor, University of Miami).

"Warmer Seas Will Wipe Out Plankton, Source of Ocean Life" (Jef Huisman, professor, University of Amsterdam).

"Climate wars threaten billions" (*A Climate of Conflict*, International Alert).

"Criminologists and police officers are now beginning to speculate that one of the hidden consequences of global warming will be an increase in street crime" (Ken Pease, professor, University College London).

"Study Says Polar Bears Could Face Extinction" (Arctic Climate Impact Assessment).

"Spiders Getting Bigger—Global Warming to Blame?" (*National Geographic*).

"The panda bear could disappear in the wild unless the pace of global warming slows" (*WWF Global*).

"Surge in fatal shark attacks blamed on global warming" (*The Guardian*).

From these we learn that if a species is warm, fuzzy, cuddly, delicious, or photogenic, it faces extinction because of global warming. But if the species bites, sticks, pesters, plagues, or eats people, global warming will cause it to thrive. That global warming is also predicted to cause an increase in food production and a surplus of clement afternoons never makes the press.

What's never noticed is that all these predictions of doom are conditional and uncertain. They are conditional on the assertion of global warming itself being true. And then each of the individual catastrophes has its own uncertainty. The total uncertainty is therefore much greater than appreciated. What makes it worse is that the predictions of doom, which assume the validity of global warming, are usually taken as proof of global warming—an argument that is exactly backward.

The Bandwagon Fallacy is also known as the Come On in the Water's Fine Fallacy, a.k.a. The Grants Are Flowing Fallacy, a.k.a. A Good Chance I'll Get Quoted in the Press Fallacy. The lesson is that scientists are no more immune to fads than are civilians.

Discussion

By the time even well-conducted studies filter down through media and finally to the public, they often bear little resemblance to the actual work. The caveats, limitations, and sober warnings against extrapolation are nowhere to be found in press reports. This is natural. The public almost by definion has not developed the instincts of researchers, which are acquired by years of patient training. Yet, quite oddly, studies that go wrong at the start, i.e. those in which authors rely one or more of the statistical fallacies as shown above, are often less distorted by the press and public than are "straight" studies. Perhaps this is because these fallacies are almost always invoked to support a bias that is shared by researchers and civilians—and, of course, by the peers who review such work.

These fallacies are only a subset of the ways research can go wrong, but they are the most popular. All fallacies are harmful, especially if the research is used to support political or cultural decisions, as in the CARB study.

These fallacies occur so freqently that their elimination is unlikely.

William M. Briggs, Ph.D., is adjunct professor of statistical science at Cornell University and a statistical consultant to many companies. Contact: matt@wmbriggs.com.

Acknowledgements: I would like to thank two anonymous reviewers and the editor for suggestions to improve this paper.

REFERENCES

1. Carter T, Ferguson MJ, Hassin RR. A single exposure to the American flag shifts support toward Republicanism up to 8 months later. *Psychol Sci* 2011;22:1011-1018.
2. Bedard P. Harvard: July 4th parades are right-wing. *US News and World Report*, Jun 30, 2011. Available at: www.usnews.com/news/blogs/washingtonwhispers/2011/06/30/harvard-july-4th-parades-are-right-wing. Accessed May 21, 2014.
3. Madestam A, Yanagizawa-Drott D. Shaping the nation: the effect of Fourth of July on political preferences and behavior in the United States. Working Paper, 2011. Available at: http://www.hks.harvard.edu/fs/dyanagi/Research/FourthOfJuly.pdf. Accessed May 21, 2011.
4. California Air Resources Board. Fine particle pollution a threat to the cardiovascular health of Californians. Press Release # 11-53, Dec 8, 2011. Available at: www.arb.ca.gov/newsrel/newsrelease.php?id=268. Accessed May 21, 2014.
5. Jerrett M. Spatiotemporal Analysis of Air Pollution and Mortality in California Based on the American Cancer Society Cohort: Final Report. Contract 06-332; 2011. Available at: http://tinyurl.com/oss4r9a. Accessed May 21, 2014.
6. Angell B. Experts work to raise radon awareness. Report. Channel 3 WSIL TV, Jun 26, 2012.
7. Bräuner EV, Andersen CE, Sørensen M, et al. Residential radon and lung cancer incidence in a Danish cohort. *Environ Res* 2012;118:130-136. doi: 10.1016/j.envres.2012.05.012.
8. Howson C, Urbach P. *Scientific Reasoning: the Bayesian Approach.* 2nd ed. Chicago, Ill.: Open Court; 1993.