Contents

Journalism 219: Not enough Statistics for Journalists

Philip B. Stark, Department of Statistics, UC Berkeley

www.stat.berkeley.edu/~stark pbstark@berkeley.edu @philipbstark

Office: 403 Evans Hall. Office hours: TBA

This version: 24 January 2016. Latest version: here

Not enough Statistics for Journalists: Overview

This short (12-hour) course presents statistical concepts crucial to understand scientific results and to report science accurately. The course is conceptual, not computational. It brings out questions that should be asked about any purported scientific finding, and helps you make sense of statistical claims. Examples will be drawn primarily from news stories I've been involved in, such as fraud in the Florida Lottery, gender bias in student teaching evaluations, pornography on the Internet and the effectiveness of content filters, and Governor Schwarzenegger's acrostic expletive veto of AB1176 (2009). We might also cover issues such as dietary salt and mortality, vaccinations, earthquake risk, Internet voting, the economic impact of climate change, forecasts of extinctions caused by climate change, GMO food crops, and the reproducibility and replicability of scientific results.

There will be weekly assignments and a term project: to track a scientific result reported in the mainstream press "upstream" and assess whether the reportage is accurate and whether the science is reliable.

For fall 2015, the course will meet on four consecutive Mondays (9/14, 21, 28, and 10/5), 3-6pm in the Library of the School of Journalism.

Administrativia

Prerequisites

- High-school algebra

Format and assessment

- four 3-hour lectures
- weekly written assignments (40% of grade)
- term project (50% of grade) Instructions here
- engaged participation in class (10% of grade)

Submitting assignments: Please submit assignments by email, to pbstark@berkeley.edu, with the subject line "J219 Assignment n," where n is the number of the assignment. I accept assignments in PDF, HTML, RTF, and MarkDown, but not in Microsoft Word.

Code of conduct; attribution of work

The high academic standard at the University of California, Berkeley, is reflected in each degree awarded. Every student is expected to maintain this high standard by ensuring that all academic work reflects unique ideas or properly attributes the ideas to the original sources.

These are some basic expectations of students with regards to academic integrity: Any work submitted should be your own individual thoughts, and should not have been submitted for credit in another course unless you have prior written permission to re-use it in this course from this instructor.

All assignments must use "proper attribution," meaning that you have identified the original source and extent or words or ideas that you reproduce or use in your assignment. This includes drafts and homework assignments! If you are unclear about expectations, ask your instructor.

Do not collaborate or work with other students on assignments or projects unless the instructor gives you permission or instruction to do so.

Disability accommodations

If you need an accommodation for a disability, if you have information your wish to share with the instructor about a medical emergency, or if you need special arrangements if the building needs to be evacuated, please inform the instructor as soon as possible.

If you are not currently listed with DSP (the Disabled Students' Program) and believe you might benefit from their support, please apply online at dsp.berkeley.edu

Main topics by week

Week 1

Reading for discussion

1. Lower Blood Pressure Guidelines Could Be 'Lifesaving,' Federal Study Says, Gina Kolata, *The New York Times*, 11 September 2015. http://http://www.nytimes.com/2015/09/12/health/blood-pressure-study.html
2. Psychologists Welcome Analysis Casting Doubt on Their Work, Benedict Careyaug, *The New York Times*, 28 August 2015. http://www.nytimes.com/2015/08/29/science/psychologists-welcome-analysis-casting-doubt-on-their-work.html

3. How reliable are psychology studies? *The New York Times*, 8 September 2015.
   http://www.nytimes.com/2015/09/08/opinion/how-reliable-are-psychology-studies.html
4. W.J. Spangler, B.S., G.R. Cosgrove, M.D., H.T. Ballantine Jr., M.D., E.H. Cassem, M.D., S.L. Rauch, M.D., A. Nierenberg, M.D., and B.H. Price, M.D., 1996. Magnetic Resonance Image-guided Cingulotomy for Intractable Psychiatric Disease, *Neurosurgery*, *38*, 1071–1078.
   http://journals.lww.com/neurosurgery/Abstract/1996/06000/Magnetic_Resonance_Image_guided_Stere
   (See also questions 10ff in http://www.stat.berkeley.edu/~stark/SticiGui/Review/selfFinal1.htm)
5. The Power of Nothing: Could studying the placebo effect change the way we think about medicine? Michael Specter, *The New Yorker*, 12 December 2011.
   http://www.newyorker.com/magazine/2011/12/12/the-power-of-nothing

Assignment Below, when I refer to "popular press," I mean "quality" popular publications such as The New York Times, The Wall Street Journal, The Washington Post, The Times of London, The Atlantic, The Economist, The New Yorker, Wired, etc. Political bent doesn't matter.

1. Reasoning and Fallacies
    1. Read SticiGui: Reasoning and fallacies.
       http://www.stat.berkeley.edu/~stark/SticiGui/Text/reasoning.htm
    2. Find three examples of informal fallacies in the popular press this week.
    3. Provide a citation for each. Point out each fallacy and describe how it plays in the story.
2. Experiments and observational studies
    1. Read SticiGui: Does Treatment Have an Effect?
       http://www.stat.berkeley.edu/~stark/SticiGui/Text/experiments.htm
    2. Find examples of reporting in the popular press this week based on:
        1. a longitudinal study
        2. a survey
        3. a randomized experiment
    3. For each example,
        1. Provide a citation.
        2. Note whether a causal inference was drawn
        3. Identify whether there was a control group; if so, explain what it was.
        4. Identify at least two things that might lead to confounding in that example.
        5. Explain how the "design" of the data collection might lead to confounding or reduce confounding.

Lecture topics

- reasoning and fallacies (reading: SticiGui: Reasoning and fallacies)
- observational studies and experiments (reading: SticiGui: Does Treatment Have an Effect?)
  - longitudinal and cross-sectional studies
  - confounding
  - controls
  - historical controls
  - randomization
  - placebos and the placebo effect

Week 2

Reading for discussion

1. *SticiGui* Probability: Philosophy and Mathematical Background
   http://www.stat.berkeley.edu/~stark/SticiGui/Text/probabilityPhilosophy.htm
2. *SticiGui* Sampling
   http://www.stat.berkeley.edu/~stark/SticiGui/Text/sampling.htm
3. A Premature SPRINT To The Finish Line, Larry Husten, *Cardio/Brief*, 14 September 2015.
   http://cardiobrief.org/2015/09/14/a-premature-sprint-to-the-finish-line/
4. Antidepressant Paxil Is Unsafe for Teenagers, New Analysis Says, Benedict Carey, *The New York Times*, 16 September 2015.
   http://www.nytimes.com/2015/09/17/health/antidepressant-paxil-is-unsafe-for-teenagers-new-analysis-says.html
5. Landmark Analysis of an Infamous Medical Study Points Out the Challenges of Research Oversight, Paul Basken, *Chronicle of Higher Education*, 17 September 2015.
   http://chronicle.com/article/Landmark-Analysis-of-an/233179/
6. Schwarzenegger "F*** You": Did Gov Send Lawmaker Obscene Message Through Acrostic Poem?, *Huffington Post*.
   http://www.huffingtonpost.com/2009/10/27/schwarzenegger-sends-lawm_n_336319.html
7. Schwarzenegger Sticks It To Assemblyman, Acrostic Style, *The Atlantic*.
   http://www.theatlantic.com/politics/archive/2009/10/schwarzenegger-sticks-it-to-assemblyman-acrostic-style/29206/
8. Stark, P.B., 2010. Null and Vetoed: "Chance Coincidence"? *Chance, 23*, 43–46.
   Preprint: http://www.stat.berkeley.edu/~stark/Preprints/acrosticVeto10.pdf
9. Mom's babies born on 8-8-08, 9-9-09, 10-10-10, Elizabeth Weise, *USA TODAY*, 14 October 2010.

http://usatoday30.usatoday.com/yourlife/parenting-family/babies/2010-10-14-Birthday14_ST_N.htm

10. Stark, P.B. and D.A. Freedman, 2003. What is the Chance of an Earthquake? in *Earthquake Science and Seismic Risk Reduction*, F. Mulargia and R.J. Geller, eds., NATO Science Series IV: Earth and Environmental Sciences, v. 32, Kluwer, Dordrecht, The Netherlands, 201–213.
Preprint:
http://www.stat.berkeley.edu/~stark/Preprints/611.pdf

Assignment Find three examples of probabilities in the popular press this week relating to something scientific (not, for instance, sports). For each, give a citation, a brief synopsis of the issue/claim, and explain what "probability" means in the context of the article. In particular, explain whether the probability results from actual randomization, from a truly stochastic phenomenon, or from a (metaphorical) model. Is it a reflection of data, of physics, of expert opinion, of modeling choices, or what? Try to assess whether the number is malleable: would different assumptions have led to different numbers? Or would it take different underlying data to yield different numbers? How "objective" is each probability?

Lecture topics

- what does "probability" mean? (reading: SticiGui: Probability: Philosophy and Mathematical Background; Stark, P.B. and D.A. Freedman, 2003)
- the ontology of probability in science
  - deliberate randomization (random samples, randomized experiments)
  - stochastic phenomena (quantum physics, thermodynamics, "noise" processes)
  - models / assumptions (mostly invented–based on metaphor)
  - how do you justify a (stochastic) model?
    - scientific basis ("truth")
    - predictive power
    - compact description of data (goodness of fit)
    - other utility?
  - case study: Schwarzenegger's acrostic veto
  - case study: chance a couple has children born 8/8/08, 9/9/09, and 10/10/10
  - case study: the chance of a major earthquake in the Bay Area by 2030
- surveys (reading: SticiGui: Sampling)
  - common survey designs
  - populations, frames, parameters, and statistics

- response rates, self-selection, nonresponse bias
- margin of error and confidence intervals
  - meaningless unless the sample is random
  - need to account for bias and nonresponse
  - can't blindly plug into formula: have to account for the sample design
- priming, wording, etc.
- semi-attached figures and red herrings
- Internet surveys
- snowball samples, sampling social networks, etc.
- case studies:
  - The Hite Report
  - student evaluations of teaching

Week 3

Reading for discussion

1. Trafimowa, D. and M. Marks, 2015. Editorial. *Basic and Applied Social Psychology*, *37*,
   DOI:10.1080/01973533.2015.1012991
   http://www.tandfonline.com/doi/abs/10.1080/01973533.2015.1012991
2. Psychology Journal bans P values: Test for reliability of results 'too easy to pass', say editors. Chris Woolston, *Nature*, 26 February 2015.
   http://www.nature.com/news/psychology-journal-bans-p-values-1.17001
3. How, and why, a journalist tricked news outlets into thinking chocolate makes you thin, Sarah Kaplan, *The Washington Post*, 28 May 2015.
   http://www.washingtonpost.com/news/morning-mix/wp/2015/05/28/how-and-why-a-journalist-tricked-news-outlets-into-thinking-chocolate-makes-you-thin/
4. How to Know Whether to Believe a Health Study, Austin Frakt, *The New York Times*, 17 August 2015.
   http://www.nytimes.com/2015/08/18/upshot/how-to-know-whether-to-believe-a-health-study.html
5. Dutch sexism study comes under fire, Martin Enserink, *Science Magazine*, 23 September 2015. http://news.sciencemag.org/scientific-community/2015/09/dutch-sexism-study-comes-under-fire
6. First brain-to-brain 'telepathy' communication via the Internet, KurzweilAI, 24 September 2015.
   http://www.kurzweilai.net/first-brain-to-brain-telepathy-communication-via-the-internet
7. Andrea Stocco, Chantel S. Prat, Darby M. Losey, Jeneva A. Cronin, Joseph Wu, Justin A. Abernethy, Rajesh P. N. Rao, 2015. Playing 20

Questions with the Mind: Collaborative Problem Solving by Humans Using a Brain-to-Brain Interface, PLoS One, DOI: 10.1371/journal.pone.0137303 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137303#pone-0137303-g004

Assignment Find an example of a margin of error or confidence interval for a survey and a claim of statistical significance (for any kind of study) in the popular press this week. For each, give a citation and brief synopsis of the issue/claim and the experiment or data source. Explain/interpret the margin of error, confidence interval, and the significance level or p-value in plain language. Assess whether the uses of the margin of error, confidence interval or test of significance are appropriate. Was there a random sample? If so, what was the sample design? Was it a simple random sample? Stratified? Cluster? Multi-stage? What was the response rate? Did the authors give you enough information to evaluate the results? What is the null hypothesis for the test of significance? Is is realistic, or a "straw man"? Where does the probability "come from" in the confidence interval or test of significance? Was there a randomized experiment, a random sample, or something else? Can you tell what formulae were used to find the margin of error, confidence interval, or the p-value? Do they fit the way the data were collected, or are they mechanical calculations using irrelevant formulae?

Lecture topics

- interpreting p-values; statistical significance and practical significance
  - null and alternative hypotheses
  - what are p-values, and why do people love to hate them?
  - "straw-man" null hypotheses; parametric and nonparametric tests
  - multiplicity, significance hunting. Voodoo correlation
  - does it matter?
  - are most scientific results wrong?
  - reproducibility and replicability
  - case study: banning p-values by *Basic and Applied Social Psychology*
  - case study: telepathy over the Internet
- interpreting confidence intervals
- meta-analysis
  - how many bad studies does it take to make a good study?
  - underlying assumptions
- causal inference
  - hypothetical counterfactuals
  - importance of randomization and intervention
  - association versus causation

- intention-to-treat analyses
- quantifying the effect of treatment
- generalizability
- are Big Data a substitute for randomization and experimental design?

Week 4

Reading for discussion

1. Health Benefits of Tea? Here's What the Evidence Says, Aaron E. Carroll, *The New York Times*, 5 October 2015. http://www.nytimes.com/2015/10/06/upshot/what-the-evidence-tells-us-about-tea.html
2. McNutt, M., 2014. Editorial: Reproducibility, *Science*, *343*, p229. DOI: 10.1126/science.1250475, https://www.sciencemag.org/content/343/6168/229
3. Huff, D., 1954. *How to Lie with Statistics*, W.W. Norton & Co., NY.
4. Lottery odds: To win, you'd have to be a loser. Lawrence Mower, *Palm Beach Post*, 28 March 2014. http://www.mypalmbeachpost.com/news/news/lottery-odds-to-win-youd-have-to-be-a-loser/nfL57
5. Richard Arratia, Skip Garibaldi, Lawrence Mower, and Philip B. Stark, 2015. Some People Have All The Luck http://arxiv.org/pdf/1503.02902.pdf
6. XKCD: *Significance*. https://xkcd.com/882/
7. Freedman, D.A., 2008. Survival analysis: a primer, *The American Statistician*, *62*, 110–119. http://amstat.tandfonline.com/doi/abs/10.1198/000313008X298439 (also in Freedman, D.A., 2010. *Statistical Models and Causal Inference: A dialog with the Social Sciences*, Cambridge University Press. D. Collier, J. Sekhon, P.B. Stark, eds.)
8. Freedman, D.A., and R. Berk, 2001. Statistical Assumptions as Empirical Commitments, http://escholarship.org/uc/item/0zj8s368#page-1 (also in Freedman, D.A., 2010. *Statistical Models and Causal Inference: A dialog with the Social Sciences*, Cambridge University Press. D. Collier, J. Sekhon, P.B. Stark, eds.)

Assignment Work on your term project.

Lecture topics

- Assumptions of common methods of analysis

  - linear regression models
  - t-tests, z-tests, ANOVA
  - logit/probit models
  - survival curves
  - Cox proportional hazards models
  - better alternatives: permutation tests
  - ...
- Case studies
  - salt and mortality
  - how many species will go extinct due to climate change?
- questions to remember
  - appropriate skepticism
  - sanity checks–too good to be true?
  - whose interests are at stake? who funded the work?
  - how did the sample or study group come to be?
  - controls? randomized assignment or random sample?
  - sources of noise, measurement error, and other uncertainties
  - confounding? possible confounders?
  - statistical v practical significance; effect size
  - how are the mathematical and real-world null hypotheses related?
  - is the null hypothesis realistic? or just a straw man?
  - multiplicity and p-hacking
  - correlation versus causation
  - science by peer review or by press release? Cold fusion, BICEP gravitational waves, etc.
  - ...

Collected Reading List:

How to read a scientific paper, Adam Ruben, *Science*, 20 January 2016. http://www.sciencemag.org/careers/2016/01/how-read-scientific-paper

The Power of "The Power Pose": Amy Cuddy's famous finding is the latest example of scientific overreach. Andrew Gelman and Kaiser Fung, *Slate*, 19 January 2015. http://www.slate.com/articles/health_and_science/science/2016/01/amy_cuddy_s_power_pose_resear

How to Know Whether to Believe a Health Study, Austin Frakt, *The New York Times*, 17 August 2015.
http://www.nytimes.com/2015/08/18/upshot/how-to-know-whether-to-believe-a-health-study.html

7.9% of species will become extinct as a result of climate change
http://www.sciencemagazinedigital.org/sciencemagazine/01_may_2015?folio=571#pg113

Chocolate and Weight Loss How, and why, a journalist tricked news outlets into thinking chocolate makes you thin, Sarah Kaplan, *The Washington Post*, 28 May 2015,
http://www.washingtonpost.com/news/morning-mix/wp/2015/05/28/how-and-why-a-journalist-tricked-news-outlets-into-thinking-chocolate-makes-you-thin/

Science by Press Release

Academics Seek a Big Splash, Noam Scheiber, *The New York Times*, 31 May 312015.
http://www.nytimes.com/2015/06/01/business/beyond-publish-or-perish-scientific-papers-look-to-make-splash.html

Lower Blood Pressure Guidelines Could Be 'Lifesaving,' Federal Study Says, Gina Kolata, *The New York Times*, 11 September 2015.
http://http://www.nytimes.com/2015/09/12/health/blood-pressure-study.html

Acrostic Veto

Schwarzenegger "F*** You": Did Gov Send Lawmaker Obscene Message Through Acrostic Poem?, *Huffington Post*,
http://www.huffingtonpost.com/2009/10/27/schwarzenegger-sends-lawm_n_336319.html

Schwarzenegger Sticks It To Assemblyman, Acrostic Style, *The Atlantic*,
http://www.theatlantic.com/politics/archive/2009/10/schwarzenegger-sticks-it-to-assemblyman-acrostic-style/29206/

Stark, P.B., 2010. Null and Vetoed: "Chance Coincidence"? *Chance, 23*(4), 43–46.
Preprint: http://www.stat.berkeley.edu/~stark/Preprints/acrosticVeto10.pdf

Student Evaluations of Teaching

Student Course Evaluations Get An 'F, Anya Kamenetz, *NPR Education Blog*, 26 September 2014. (Evaluating teaching, misuse of Statistics)
http://www.npr.org/blogs/ed/2014/09/26/345515451/student-course-evaluations-get-an-f

2 scholars flunk course evaluations as measures of teaching quality, Dan Berrett, *Chronicle of Higher Education*, p. A16, 26 September 2014.
http://chronicle.texterity.com/chronicle/20140926a?sub_id=2FQNKVDXMnsU#pg16

Stark, P.B., and R. Freishtat, 2014. An evaluation of course evaluations. *Science Open*, DOI 10.14293/S2199-1006.1.-.AOFRQA.v1,
https://www.scienceopen.com/document/vid/42e6aae5-246b-4900-8015-dc99b467b6e4

McNell, L., A. Driscoll, and A.N. Hunt, 2014. What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*,
http://link.springer.com/article/10.1007%2Fs10755-014-9313-4#page-1

Fraud in the Florida Lottery

Lottery odds: To win, you'd have to be a loser. Lawrence Mower, *Palm Beach Post*, 28 March 2014.
http://www.mypalmbeachpost.com/news/news/lottery-odds-to-win-youd-have-to-be-a-loser/nfL57

Richard Arratia, Skip Garibaldi, Lawrence Mower, and Philip B. Stark, 2015. Some People Have All The Luck
http://arxiv.org/pdf/1503.02902.pdf

Porn on the Internet

Internet Content Filters Fail to Block Sexually Explicit Material. Thomas Claburn, *Information Week*, 14 November 2006.
http://www.informationweek.com/news/showArticle.jhtml?articleID=194300677&section=All+Stories

1 percent of Web sites deemed pornographic. Maryclaire Dale, *Associated Press*, 14 November 2006.
http://www.msnbc.msn.com/id/15721799/

U.S., Google Set to Face Off in Court. Michael Liedtke, *Associated Press*, 14 March 2006.
http://www.sfgate.com/cgi-bin/article.cgi?file=/n/a/2006/03/13/financial/f133050S47.DTL&type=printable

Google privacy issue enters court. *Al Jazeera*, 14 March 2006.
http://english.aljazeera.net/archive/2006/03/200841013165547373 7.html

In Case About Google's Secrets, Yours Are Safe. Adam Liptak, *The New York Times*, 26 January 2006.
http://www.nytimes.com/2006/01/26/technology/26privacy.html?_r=1&emc=eta1&oref=slogin

Google Resists U.S. Subpoena of Search Data. Katie Hafner and Matt Richtel, *The New York Times*, 20 January 2006.
http://www.nytimes.com/2006/01/20/technology/20google.html?pagewanted=1&_r=1

Feds take porn fight to Google. Declan McCullagh and Elinor Mills, *CNET News*, 19 January 2006.
http://news.cnet.com/Feds-take-porn-fight-to-Google/2100-1030_3-6028701.html?tag=mncol;txt

Stark, P.B., 2008. The effectiveness of Internet content filters. *I/S: A Journal of Law and Policy for the Information Society, 4*, 411–429.
Reprint: http://www.is-journal.org/V04I02/Stark.pdf
Preprint: http://www.stat.berkeley.edu/~stark/Preprints/filter07.pdf

Earthquake probabilities

USGS 2008 Bay Area Earthquake Probabilities.
http://earthquake.usgs.gov/regional/nca/ucerf/

Stark, P.B. and D.A. Freedman, 2003. What is the Chance of an Earthquake? in *Earthquake Science and Seismic Risk Reduction*, F. Mulargia and R.J. Geller, eds., NATO Science Series IV: Earth and Environmental Sciences, v. 32, Kluwer, Dordrecht, The Netherlands, 201–213.
Preprint: http://www.stat.berkeley.edu/~stark/Preprints/611.pdf

Gravitational waves

Gravitational Waves from Big Bang Detected: A curved signature in the cosmic microwave background light provides proof of inflation and spacetime ripples Clara Moskowitz, *Scientific American*, 17 March 2014.
http://www.scientificamerican.com/article/gravity-waves-cmb-b-mode-polarization/

Gravitational wave evidence disappears into dust There may still be evidence of cosmic inflation, but it will be hard to spot. John Timmer, *Ars Technica*, 22 September 2014.
http://arstechnica.com/science/2014/09/22/gravity-wave-evidence-disappears-into-dust/

Economic Impact of Climate Change

The American Climate Prospectus: Economic Risks in the United States, 2014. Trevor Houser, Robert Kopp, Solomon Hsiang, Michael Delgado, Amir Jina, Kate Larsen, Michael Mastrandrea, Shashank Mohan, Robert Muir-Wood, DJ Rasmussen, James Rising, and Paul Wilson.
http://rhg.com/reports/climate-prospectus

Saltelli, A., P.B. Stark, W. Becker, and P. Stano, 2015. Climate Models as Economic Guides: Scientific Challenge or Quixotic Quest?, *Issues in Science and Technology*, Spring 2015.
Reprint: http://www.stat.berkeley.edu/~stark/Preprints/saltelliEtal15.pdf

The p-value Ban

Psychology journal bans P values: Test for reliability of results 'too easy to pass', say editors. Chris Woolston, *Nature*, 26 February 2015.
http://www.nature.com/news/psychology-journal-bans-p-values-1.17001

Introduction to Statistics

Stark, P.B., 1997-2015. *SticiGui*, http://www.stat.berkeley.edu/~stark/SticiGui

Huff, D., 1954. *How to Lie with Statistics*, Norton.