

Lectures on Foundations of Statistics and Inference

P.B. Stark www.stat.berkeley.edu/~stark

- **This document:** www.github.com/pbstark/Kavli18/kavliStat18.pdf
- **Repository for the course:** www.github.com/pbstark/Kavli18/

Last edited 10 July 2018, 5:03GMT

Reading for the week, and other references

1. Boring, A., K. Ottoboni, and P.B. Stark, 2016. Student evaluations of teaching (mostly) do not measure teaching effectiveness, *ScienceOpen Research*, DOI: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1 <https://www.scienceopen.com/document?vid=818d8ec0-5908-47d8-86b4-5dc38f04b23e>
2. Cornell, C.A., 1968. Engineering seismic risk analysis, *Bull. Seism. Soc. Am*, 58, 1583–1606.
3. Feynman, R., 1974. CalTech Commencement Address, <http://calteches.library.caltech.edu/51/2/CargoCult>
4. Freedman, D.A., 1995. Some issues in the foundations of statistics, *Foundations of Science*, 1, 19–39. <https://doi.org/10.1007/BF00208723>
5. Freedman, D.A., 1999. From association to causation: some remarks on the history of statistics, *Statistical Science*, 14(3), 243–258.
6. Freedman, D.A., 2008. On types of scientific inquiry: the role of qualitative reasoning, *The Oxford Handbook of Political Methodology*, Box-Steffensmeier, J.M., H.E. Brady, and D. Collier (eds), Oxford University Press, Oxford. DOI: 10.1093/oxfordhb/9780199286546.003.0012. [Preprint](#)
7. Freedman, D.A., 2009. *Statistical Models: Theory and Practice*, 2nd edition, Cambridge University Press.
8. Freedman, D.A., R. Pisani, and R. Purves, 2007. *Statistics, 4th edition*, W.W. Norton, New York.
9. Klemes, V., 1989. The Improbable Probabilities of Extreme Floods and Droughts, in O. Starosolsky and O.M. Meldev (eds), *Hydrology and Disasters*, James and James, London, 43–51. https://www.itia.ntua.gr/en/getfile/1107/1/documents/1997_ImprobProbabilities_OCR.pdf
10. LeCam, L., 1977. Note on metastatistics or ‘An essay toward stating a problem in the doctrine of chances,’ *Synthese*, 36, 133–160.
11. Morabia, A., 2006. Pierre-Charles-Alexandre Louis and the evaluation of bloodletting, *J. Roy. Soc. Medicine*, 99, 158–160. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1383766/pdf/0158.pdf>

12. Mulargia, F., P.B. Stark, and R.J. Geller, 2017. Why is probabilistic seismic hazard analysis (PSHA) still used?
13. 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. <https://www.stat.berkeley.edu/~stark/Preprints/611.pdf>
14. Stark, P.B., 1997. [SticiGui](#) Various chapters assigned below.
15. Stark, P.B., 2016a. [Pay no attention to the model behind the curtain](#)
16. Stark, P.B., 2016b. The value of P-values, *The American Statistician*, 70, DOI:10.1080/00031305.2016.1154108
17. Stark, P.B., 2017. [Mathematical Foundations](#), [Inequalities](#), [Introduction to permutation tests](#), [Rabbits and Cargo-Cult Statistics](#), [Generating pseudo-random samples and permutations](#)
18. Stark, P.B., and A. Saltelli, 2018. Cargo-cult Statistics and Scientific Crisis, *Significance*, 15(4), 40–43. Preprint: <https://www.significancemagazine.com/2-uncategorised/593-cargo-cult-statistics-and-scientific-crisis>
19. Urban, M.C., 2015. Accelerating extinction risk from climate change, *Science*, 348, Issue 6234, 571–573, DOI: 10.1126/science.aaa4984, <http://science.sciencemag.org/content/348/6234/571.full>

Lecture 1: Introduction to Inference.

Importance of subject-matter knowledge, critical thinking, and curiosity. The method of comparison, experiments and observational studies, Snow's investigation of the cause of cholera, Pierre Louis and bloodletting, confounding, randomization, controls, blinding, the Neyman model for causal inference, the Lady Tasting Tea experiment, Fisher's Exact Test, response schedules.

Assignment.

1. Read:
 - SticiGui [Counting](#)
 - SticiGui [Theories of Probability](#)
 - SticiGui [Naive set theory](#)
 - SticiGui [Logic](#)
 - SticiGui [Axioms of Probability](#)
 - SticiGui [Experiments](#)
 - Freedman (2008)
 - Morabia (2006)
2. Work the self-test problems in the chapters of SticiGui assigned above.
3. **To hand in:** Look at [the data Morabia transcribed from P.C.A. Louis on bloodletting for pneumonia](#). Is Louis's work an observational study or

an experiment? Do you think it amounts to a “natural experiment,” like Snow’s work on Cholera? Why or why not? Give two scientific questions (*statistical hypotheses*) those data might address. What do you think the most important confounding factors would be, for those two hypotheses? What would be the most natural “as-if” randomization to use in analyzing the data to address the hypotheses you formulated, if you were to consider the data to be a natural experiment? What are the controls? Is the experiment blind? Double-blind? Explain how you might use the Neyman model to analyze the data. Is the randomization assumption reasonable? Why or why not? Is the non-interference assumption reasonable? Why or why not?

Notes.

- No amount of computation or theory can compensate for bad study design—but it can hide the issues by changing the subject.
- To answer real-world scientific questions, critical thinking and subject-matter knowledge are more important than mathematical, statistical, or computational technique.

Wisdom.

There is a natural desire to substitute intellectual capital for labor, and an equally-natural preference for system and rigor over methods that seem more haphazard.

– David A. Freedman, 2008

Naturally, there is a strong desire to substitute intellectual capital for labor. That is why investigators often try to base causal inference on statistical models. With this approach, P-values play a crucial role. The technology is relatively easy to use and promises to open a wide variety of questions to the research effort. However, the appearance of methodological rigor can be deceptive. Like confidence intervals, P-values generally deal with the problem of sampling error not the problem of bias. Even with sampling error, artifactual results are likely if there is any kind of search over possible specifications for a model, or different definitions of exposure and disease. Models may be used in efforts to adjust for confounding and other sources of bias, but many somewhat arbitrary choices are made. Which variables to enter in the equation? What functional form to use? What assumptions to make about error terms? These choices are seldom dictated either by data or prior scientific knowledge. That is why judgment is so critical, the opportunity for error so large and

the number of successful applications so limited.

– David A. Freedman, 1999

If we are uncritical we shall always find what we want: we shall look for, and find, confirmations, and we shall look away from, and not see, whatever might be dangerous to our pet theories. In this way it is only too easy to obtain what appears to be overwhelming evidence in favor of a theory which, if approached critically, would have been refuted.

– Karl Popper

Whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve.

– Karl Popper

Fallacies do not cease to be fallacies because they become fashions.

– G.K. Chesterton

Snow on London cholera epidemic of 1853-1854.

John Snow was a nineteenth-century physician in London. In 1855, decades before the germ theory of disease was accepted, Snow showed that cholera is caused by an infectious organism that lives in water. His argument had many facets: an apparent time lag between infection and symptoms, explained by the time it takes the organism to reproduce in the human body; propagation of the disease along trade routes; sailors coming from places free of cholera, then visiting ports where there was cholera, did not get sick until they came in contact with locals; identifying the first and second cases in the 1848 London cholera epidemic (the first was a seaman named John Harnold who had just come from Hamburg, Germany, where there was a cholera outbreak; the second was the person who stayed in the room Harnold had used, after Harnold died). Snow found apartment buildings where many people had died, adjacent to apartment buildings where few or none had died; their water suppliers differed. Following an outbreak of cholera in 1854, Snow made a map of the residences of victims. They were concentrated near a public water pump, the Broad Street pump in Soho. A few buildings in the area were relatively unaffected by cholera; it turned out that their water supplies were different (a brewery and a poorhouse, both of which had their own pumps). Snow showed that most of the cholera victims in other parts of London had drunk from the Broad Street pump.

“Miasma” was the competing theory. William Farr, prominent medical statistician at the time, claimed pollution (noxious smells, etc.) were the cause. Farr proposed the model $a/(b + x)$, where x is elevation. It fit the data for the 1848-1849 epidemic very well.

In Snow’s words:

Although the above facts shown in the table above afford very strong evidence of the powerful influence which the drinking of water containing the sewage of a town exerts over the spread of cholera, when that disease is present, yet the question does not end here; for the intermixing of the water supply of the Southwark and Vauxhall Company with that of the Lambeth Company, over an extensive part of London, admitted of the subject being sifted in such a way as to yield the most incontrovertible proof on one side or the other. In the subdistricts enumerated in the above table as being supplied by both Companies, the mixing of the supply is of the most intimate kind. The pipes of each company go down all the streets, and into nearly all the courts and alleys. A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time when the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different Companies. Now it must be evident that, if the diminution of cholera, in the districts partly supplied with improved water, depended on this supply, the houses receiving it would be the houses enjoying the whole benefit of the diminutions of the malady, whilst the houses supplied by the water from Battersea Fields would suffer the same mortality as they would if the improved supply did not exist at all. As there is no difference whatever in the houses or the people receiving the supply of the two Water Companies, or in any of the physical conditions with which they are surrounded, it is obvious that no experiment could have been devised which would more thoroughly test the effect of water supply on the progress of cholera than this, which circumstances placed ready made before the observer.

The experiment, too, was on the grandest scale. No fewer than three hundred thousand people of both sexes, of every age and occupation, and of every rank and station, from gentlefolks down to the very poor, were divided into groups without their choice, and in most cases, without their knowledge; one group being supplied with water containing the sewage of London, and amongst it, whatever might have come from the cholera patients; the other group having water quite free from such impurity.

To turn this grand experiment to account, all that was required was to learn the supply of water to each individual house where a fatal attack of cholera might occur.

Lambeth water company started drawing its water further upstream in 1852 (water is cleaner upstream of the city, because refuse and sewage are dumped

into the river as the river flows through the city). This allowed Snow to compare the rates of cholera in the 1853–1854 epidemics (in which about 2,800 people died, more than 500 in a single 10-day period) with earlier epidemics, when the Lambeth company drew its water further downstream, along with one of its competitors, the Southwark and Vauxhall company. Which buildings were served by which water company was largely accidental: other than water supplier, there was not much difference that could account for the differences in the rate of cholera.

Cholera deaths by water source, London epidemic of 1853–1854. Snow’s Table IX, via Freedman, 1999.

water supplier	houses	deaths from cholera	deaths per 10,000
Southwark & Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
rest of London	256,423	1,422	59

Louis and bloodletting

Pierre-Charles-Alexandre Louis was an early 19th century physician in Paris. At the time, bloodletting (by leaches or lancets) was a commonly accepted remedy for almost anything; the theory was that illnesses were caused by inflammation of various organs, and that bleeding the subject reduced that inflammation. The leading physician promoting the use of leeches, Broussais, was so successful that France imported 42 million leeches in 1833.

Louis compared pneumonia patients bled “early” with those bled “late.”

Lecture 2: Probability Models

Equally likely outcomes, the frequency theory, the subjective theory, probability as metaphor, probability models, Kolmogorov’s axioms, consequences of the axioms, partitions, conditional probability, the multiplication rule, Bayes’ rule, the law of total probability, useful probability inequalities. Box models, sampling distributions, the Law of Large Numbers, the expected value, common distributions arising from 0-1 boxes. Parameters, statistics, and estimation. Ontology of probability in applications. Bayesian and frequentist uncertainties; credible intervals and confidence intervals. Probability models in practice: earthquake probabilities, climate change and extinctions, birds and wind turbines.

Assignment

1. Read:
 - SticiGui [Let's Make a Deal](#)
 - SticiGui [Probability Meets Data](#)
 - SticiGui [Random Variables](#)
 - SticiGui [Expectation](#)
 - [Mathematical Foundations](#)
 - [Inequalities](#)
 - Freedman (1995)
 - [Rabbits and Cargo-Cult Statistics](#)
 - Klemes (1989)
 - LeCam (1977)
 - Stark and Freedman (2003)
 - Mulargia et al. (2017)
 - Urban (2015), including [the supplementary materials](#)
 2. Work the self-test problems in the chapters of SticiGui assigned for this lecture.
 3. Mathematics students, please find a non-mathematics student, and vice versa. Discuss the LeCam, Klemes, and Urban papers with each other; explain to each other (or figure out together) any concepts you don't understand.
 4. **To hand in:** Urban estimates that 7.9% of species will become extinct as a result of climate change, with a 95% confidence interval of [6.2%, 9.8%]. Explain how Urban calculates the estimate and uncertainty. Is his work based on a random sample of species? Of geography? Explain his statistical model. Identify where in his work he assumes that associations are response schedules. (Hint: one of his references is McDonald, K.A., and J.H. Brown, 1992. Using montane mammals to model extinctions due to global change. *Conserv. Biol.*, 6, 409–415. doi:10.1046/j.1523-1739.1992.06030409.x) For each such assumption, say whether you think it is reasonable, and explain why. Some aspects of the paper are Bayesian. What are the priors? What do the uncertainties mean? Is the interval [6.2%, 9.8%] a confidence interval? List 5 sources of uncertainty that he omits. Do you think those sources are smaller or larger than those he includes? Explain.
-

Lecture 3: Tests

Statistical hypotheses, null and alternative hypotheses, hypothesis tests, families of tests, P-values, the meaning of P-values, abuses of P-values, group invariances of the distribution under the null hypothesis, conditional hypothesis tests, test statistics, Application: gender bias in teaching evaluations.

Assignment

1. Read:
 - Stark, 2016b.
 - Stark and Saltelli, 2018.
 - [Introduction to permutation tests](#)
 - [Generating pseudo-random samples and permutations](#)
 - Boring, Ottoboni, and Stark, 2016.
2. **To hand in:** Implement permutation tests (in R or Python) for the hypotheses about bloodletting you formulated for the first assignment, using the data Morabia transcribed from Louis. Discuss numerical issues in implementing the permutation tests. What pseudo-random number generator did you use? How did you set the seed? How did you choose the number of random permutations to perform? (Is there a principle you might use to decide the number? If so, what?) What test statistic(s) did you use? Why did you pick those? How to they connect to the relevant alternative hypotheses? Do you think the results would be qualitatively different if you had used a different test statistic? Discuss your findings, including any weaknesses in your framing of the problem (the statistical assumptions) and in the numerical analysis.