2020

Clear-Sighted Statistics: Module 19: Wrapping Up (slides)

Edward Volchok
CUNY Queensborough Community College

Wrapping Up

Module 19

CLEAR-SIGHTED STATISTICS

EDWARD VOLCHOK

0

"The first principle is that you must not fool yourself—and you are the easiest person to fool."

-- Richard P. Feynman

Richard P. Feynman, "Cargo Cult Science: Some Remarks on Science Pseudoscience, and Learning How Not to Fool Yourself," California Institute of Technology 1974 Commencement Address

1

"One of the most frustrating aspects of the journal business is the null hypothesis. It just will not go away.... It is impossible to drag authors away from their p values, and the more zeros after the decimal point, the harder people cling to them. It is almost as if all the statistics courses in the world stopped after introducing Type I error....Perhaps p values are like mosquitoes. They have an evolutionary niche somewhere and no amount of scratching, swatting, or spraying will dislodge them....investigators must learn to argue for the [practical] significance of their results without reference to inferential statistics."

-- John P. Campbell
Editor, Journal of Applied Psychology
"Some Remarks From the Outgoing Editor"
1982

John P. Campbell, "Editorial: Some Remarks From the Outgoing Editor," *Journal of Applied Psychology*, Vol. 67., No. 6, 1982, p. 698).

2

"Sir Ronald [Fisher] has befuddled us, mesmerized us, and led us down the primrose path. I believe that the almost universal reliance on merely refuting the null hypothesis as the standard method for corroborating substantive theories in the soft areas [personality and social psychology] is a terrible mistake, is basically unsound, poor scientific strategy, and one of the worse things that ever happened in the history of psychology."

-- R. Chris Fraley
Psychologist, University of Illinois at Urbana-Champaign
Cited in *The Cult of Statistical Significance*
Pp. 128-9

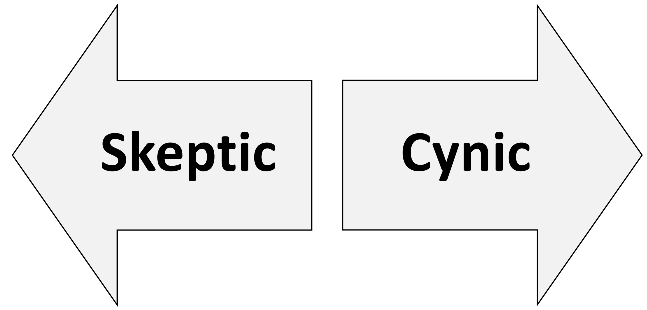
Stephen T. Ziliak and Deirdre N. McCloskey, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, (Ann Arbor, MI: The University of Michigan Press, 2008)

3

Key takeaway from Intro. to Statistics

- Facts based on sample statistics are probabilistic
- We do not have 100 percent certainty
- We never consider the H_0 true when we fail to reject it
- H_1 is not true when the H_0 is rejected

Be a Skeptic, not a Cynic



Cathy O’Neil on being a skeptic*

“A skeptic is someone who maintains a consistently inquisitive attitude toward facts, opinions, or (especially) beliefs stated as facts. A skeptic asks questions when confronted with a claim that has been taken for granted. That’s not to say a skeptic brow-beats someone for their beliefs, but rather that they set up reasonable experiments to test those beliefs.”

*Cathy O’Neil, *On Being a Data Skeptic*. (Cambridge: O'Reilly Media, 2013). Kindle Edition. Location.
O’Neil is author of *Weapons of Math Destruction*, (New York: Broadway Books, 2017).

What is covered in more advanced courses?

Clear-Sighted Statistics

Time Series Analysis

8

8

Clear-Sighted Statistics

Secular Trends

- Long-term non-periodic variation in the longitudinal data
- The timescale used is a key determinant on whether longitudinal data are perceived as a secular trend
- The aging of the population of an advanced post-industrial country
- Expansion of digital technologies
- The reliance on fossil fuels like coal, oil, and natural gas
- Trends in global warming

9

9

Clear-Sighted Statistics

Cyclical Variation

- Oscillating movements in time series data
- The business cycle's swings between boom and bust are a classic example of cyclical variations

10

10

Clear-Sighted Statistics

Seasonal Variation

- Repeated changes in time series data within a year
- Ice cream sales on the Coney Island boardwalk
- Number of people employed at ski resorts
- Christmas trees sales

11

11

Random Variation

Variations in the time series data that do not follow a predictable model

Impact on the American economy of impeaching the President of the United States

Decision Theory

Decision Theory

Evaluation of decision-makers' choices based on the possible outcome

Statistical information informs decision-makers of the uncertainties—the probabilities—involved in the decision

Decision theory is a major topic in graduate-level management curriculum

Statistical Process Control (SPC)

Statistical Process Control, SPC

Techniques used to improve the quality of manufacturing processes

Six Sigma (6σ): Data-driven process to reduce manufacturing errors to ≥ 3.4 out of a million production units

The name, *Six Sigma*, comes the error goal of one in 3.4 million would be six standard deviations from the mean

Meta-Analysis

Meta-Analysis

Collection of quantitative procedures that synthesizes the findings from a literature review of research of the topic under investigation

More Advanced ANOVA

Advanced ANOVA Tests

- One-Way Repeated Measures ANOVA
- Two-Way Anova Without Replication
- Two-Way Anova With Replication
- Factorial ANOVA
- MANOVA
- ANCOVA
- Kruskal-Wallis H

More Advanced Regression

Advance Regression Techniques

- Multiple Regression
- Spearman's Rho
- Kendall's Tau
- Logistic Regression
- Multinomial Logistic Regression
- Structural Equation Modeling

Nonparametric Techniques

Nonparametric Techniques

Parametric Test	Nonparametric Tests
One-sample z-test, One-sample t-test	Sign test
One-sample z-test, One-sample t-test	Wilcoxon Signed Rank test
Two-sample t-test for independent means	Wilcoxon-Mann-Whitney test
One-way ANOVA test	Kruskal-Wallis test and Mood's Median test
Two-way ANOVA test	Friedman test
Coefficient of Correlation	Spearman Rank Correlation

24

Bayesian Inference

25

Bayesian Inference vs. NHST

NHST only provides evidence against the plausibility of the H_0, but fails to provide any evidence in favor of the H_1

Inferences are made on hypothetical data distributions (z-, t-, F-, or chi-square, among others) instead of being based on actual data

No clear rules for stopping data collection and as a result any H_0 can be rejected when the sample is large enough

26

What Bayesian Inference does

Degree of belief in the H_0 & H_1—our "prior knowledge"—is updated in light of new data

Researchers review magnitude of evidence that supports the existence of an effect, instead of a dichotomous decision that an effect either exists or does not exist

27

Three sources of information

A model that specifies how latent parameters (ϕ) generate data (D)

Prior information about those parameters

The observed data (likelihood)

Bayes Factors: Lend support to either the H_0 or H_1

$$\text{Bayes Factors} = \frac{P(\text{Data} \mid H_1)}{P(\text{Data} \mid H_0)}$$

Interpreting Bays Factors

Bayes Factor	Interpretation
> 100	Extreme evidence for the Alternate Hypothesis
30 – 100	Very strong evidence for the Alternate Hypothesis
10 - 30	Strong evidence for the Alternate Hypothesis
3 - 10	Moderate evidence for the Alternate Hypothesis
1 - 3	Anecdotal evidence for the Alternate Hypothesis
1	No evidence
1/3 - 2	Anecdotal evidence for the Null Hypothesis
1/3 - 1/10	Moderate evidence for the Null Hypothesis
1/10 - 1/30	Strong evidence for the Null Hypothesis

Is Statistics in the Midst of a Paradigm Shift?

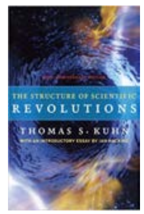
A Paradigm Shift

Term coined by Kuhn in the *Structure of Scientific Revolutions* (1962)

Paradigm shift: A fundamental change in accepted scientific thinking and methods

Paradigm shifts have four stages

Thomas S. Kuhn
(1922 – 1996)



32

Stage #1: Normal Science

Dominant paradigm defines how science is conducted

Dominant paradigm is active and widely supported

33

#2: Extraordinary Research

The dominant paradigm becomes suspect when researchers find anomalies

"Confronted with anomaly or with crisis, scientists take a different attitude toward existing paradigms, and the nature of research changes accordingly."*

*Thomas S. Kuhn, *The Structure of Scientific Revolutions*, (Chicago, IL: University of Chicago Press, 2012). P. 91.

34

#3: Adoption of a New Paradigm

Scientists conducting extraordinary research eventually develop a new paradigm

Many scientists refuse to adopt the new paradigm

*Thomas S. Kuhn, *The Structure of Scientific Revolutions*, (Chicago, IL: University of Chicago Press, 2012). P. 150.

35

#3: Adoption of a New Paradigm

"…a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents die, and a new generation grows up that is familiar with it."*

-- Max Planck quoted by Kuhn

Max Planck
(1858 – 1947)

*Thomas S. Kuhn, *The Structure of Scientific Revolutions*, (Chicago, IL: University of Chicago Press, 2012). P. 150.

#4: Aftermath

New paradigm becomes dominant

New textbooks are written

Is Statistics in Stage 3?

Is NHST on its way to the dustbin of history?

Long history of arguments against NHST

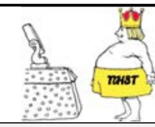
Criticism of the NHST paradigm go back at least as far as 1951

1951: Frank Yates (An associate of R. A. Fisher)

Fisher's *Statistical Methods for Research Workers* has caused researchers to "…pay undue attention to the results of the tests of significance they perform on their data…and too little to the estimates of the magnitude of the effects they are estimating."*

*Frank Yates, "The Influence of 'Statistical Methods for Research Workers' on the Development of the Science of Statistics," *Journal of the American Statistical Association*, Vol. 46, No. 253, March 1951, p. 32.

1966: The NHST Emperor has no clothes



"…***the test of significance*** does not provide the information concerning psychological phenomena characteristically attributed to it; and that, furthermore, a ***great deal of mischief*** has been associated with its use. ***What is said in this paper is hardly original.*** It is, in a certain sense, what '***everybody knows***.' To say it 'out loud' is, as it were, to assume the role of the child who pointed out that the ***emperor was really outfitted in his underwear***."* (Italics added)

-- David Bakan, Psychologist & Bayesian Statistician

*David Bakan, "Tests of Significance in Psychological Research," *Psychological Bulletin*, Vol. 66, December 1966, p. 423.

1970: Significance Test Controversy

"[NHST is] a potent but sterile intellectual rake [a shamelessly immoral person] who leaves in his merry path a long train of ravished maidens but no viable scientific offspring."*

*Paul Meehl's "Theory Testing in Psychology ad Physics: A Methodological Paradox," in Ramon E. Morrison and Denton E. Henkel, *Significance Test Controversy*, (Chicago, IL, Butterworths, 1970), p. 265.

Long-Term Critic Jacob Cohen

"After 4 decades of severe criticism, the ritual of Null Hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists."*

*Jacob Cohen, "The Earth is Round (p <.05)," Vol. 49, No. 12. *American Psychologist*, December, 1994, p. 997.

Cohen's 3 concerns about NHST

Misinterpretation: p-values as the probability that the H_0 is false

Misinterpretation: Complement of p-values is the probability of successful replication of the study

Mistaken assumption that if one rejects H_0, the theory that led to the test is affirmed

Jacob Cohen, "The Earth is Round (p < .05)," Vol. 49, No. 12. *American Psychologist*, December, 1994, p. 997.

2005:


Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

Modeling the Framework for False Positive Findings

It can be proven that most claimed research findings are false.

2005: "Most Published Research Findings are False"

"There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field."*

-- John P. A. Ioannidis

*John P. A. Ioannidis, "What Most Published Research Findings are False," *PLoS Med*, August 2005 Vol. 2(8) e.124.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/

Ioannidis' Six Corollaries (#1 to #3)

The smaller the sample sizes, the less likely the research findings are to be true

The smaller the effect sizes in a scientific field, the less likely the research findings are to be true

The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true

*John P. A. Ioannidis, "What Most Published Research Findings are False," *PLoS Med*, August 2005 Vol. 2(8) e.124.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/

Ioannidis' Six Corollaries (#4 - #6)

The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true

The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true

The hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true

*John P. A. Ioannidis, "What Most Published Research Findings are False," *PLoS Med*, August 2005 Vol. 2(8) e.124. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/

American Statistical Society (ASA) takes aim at NHST

2016: ASA's Statement on p-values and statistical significance

"Statisticians and others have been sounding the alarm about these matters for decades, to little avail. We hoped that a statement from the world's largest professional association of statisticians [the American Statistical Association] would open a fresh discussion and draw renewed and vigorous attention to changing the practice of science with regards to the use of statistical inference."*

*Ronald L. Wasserstein and Nicole A. Lazar, "The ASA Statement on p-Values: Context Process, and Purpose," *The American Statistician*, Vol. 70, No. 20, March 7, 2016, p. 130. https://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

2016: ASA's Six Points (#1 - #3)

p-values indicate how incompatible the data are with a specified statistical model

p-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone

Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold

"The ASA Statement on *p*-Values: Context Process, and Purpose", *The American Statistician*, Vol. 70, No. 20, March 7, 2016, pp. 131-2. https://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

2016: ASA's Six Points (#4 - #6)

Proper inference requires full reporting and transparency

A p-value, or statistical significance, does not measure the size of an effect or the importance of a result

By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis

"The ASA Statement on *p*-Values: Context Process, and Purpose", *The American Statistician*, Vol. 70, No. 20, March 7, 2016, pp. 131-2.
https://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

Clear-Sighted Statistics

Summarizing the 2016 Statement

"…data analysis should not end with the calculation of a p-value when other approaches are appropriate and feasible"

Approaches include: "…confidence, credibility, or prediction intervals; Bayesian methods; alternative measures of evidence, such as likelihood ratios or Bayes Factors; and other approaches such as decision-theoretic modeling and false discovery rates."

"The ASA Statement on *p*-Values: Context Process, and Purpose", *The American Statistician*, Vol. 70, No. 20, March 7, 2016, pp. 131-2.
https://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

CC BY NC SA

Clear-Sighted Statistics

2019: ASA Gets Tougher

"The *ASA Statement on P-Values and Statistical Significance* stopped just short [in 2016] of recommending that declarations of 'statistical significance' be abandoned. We take that step here. We conclude, based on our review of the [43] articles in this special issue and the broader literature, that it is time to stop using the term 'statistically significant' entirely. Nor should variants such as 'significantly different,' '*p* < 0.05,' and 'nonsignificant' survive, whether expressed in words, by asterisks in a table, or in some other way."*

*Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar (2019) Moving to a World Beyond "p < 0.05", The American Statistician, 73:sup1, p. 2.
https://www.tandfonline.com/doi/pdf/10.1080/00031305.2019.1583913?needAccess=true

CC BY NC SA

Clear-Sighted Statistics

43 papers followed the editorial

"[these papers] do not sing as one. At times in this editorial and the papers you'll hear deep dissonance, the echoes of 'statistics wars' still simmering today."*

*Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar (2019) Moving to a World Beyond "p < 0.05", The American Statistician, 73:sup1, p. 2.
https://www.tandfonline.com/doi/pdf/10.1080/00031305.2019.1583913?needAccess=true

CC BY NC SA

Inferential Statistics is entering Stage 3 of a Paradigm Shift

"It's tough to make predictions, especially about the future"* -- Yogi Berra

Here is what might happen…

*Yogi Berra, *Good Reads*, https://www.goodreads.com/quotes/261863-it-s-tough-to-make-predictions-especially-about-the-future

Expect greater emphasis on:

- Confidence Intervals
- Effect Size
- Statistical Power
- Reproducible Results
- Meta-Analysis

The term statistically significant will be deemphasized, if not banished

CLEAR-SIGHTED STATISTICS

EDWARD VOLCHOK