

The Ethical Practice of Statistics (for the Perplexed): A Course in the Social and Behavioral Sciences

Lawrence Hubert

Department of Psychology
The University of Illinois

Joint Statistical Meetings Presentation
Montreal, Canada
August, 2013

Catalog Title: Ethical Practice of Statistics

The Ethical Practice of Statistics (for the Perplexed): A Course in the Social and Behavioral Sciences

Lawrence Hubert

The course I will be discussing actually has the longer title:
Psychology (and Statistics) 484: Statistics, Ethics, and the Social and Behavioral Sciences

It will be offered for the first time at Illinois in the Spring Semester of 2014, jointly between the Departments of Psychology and Statistics.

It is directed toward upper-division undergraduates (particularly those who are pre-law, pre-med, or majoring in statistics); also, first-year graduate students in statistics or any of the social and behavioral sciences.

Course Synopsis: Course Catalog Entry

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

This course is devoted to the ethical practice of statistics, defined as being in accord with the accepted rules and standards for right conduct that govern the discipline of statistics and its many areas of application.

An emphasis is placed on the use of statistical and probabilistic reasoning in the social, behavioral, and biomedical sciences, with particular stress on the relation to law and the judiciary.

Among others, topics include the use of Bayes Rule in screening for rare events;
the importance of baserates;
probability of causation;
inferring causality;
meta-analysis;
ethics in data collection;
the Federal Rules of Evidence;
experimental design and data interpretation;
Simpson's Paradox;
statistical sleuthing and explanation;
observational and controlled studies;
the social construction of statistics;
prediction and correlation.

Course Materials

The Ethical Practice of Statistics (for the Perplexed): A Course in the Social and Behavioral Sciences

Lawrence Hubert

The primary text for the class is *A Statistical Guide for the Ethically Perplexed* (Acronym: SGEP), Lawrence Hubert and Howard Wainer, Chapman & Hall, 2013.

This text is available as an e-book through the Illinois library (under a CIC (Committee on Institutional Cooperation) publication agreement with Taylor and Francis; but password protected).

<http://hdl.handle.net/2142/42035>

Here is a link to the publisher's web site:

<http://www.crcpress.com/product/isbn/9781439873687>

The thirty (required) longer readings from the popular literature (e.g., New Yorker, New York Review of Books) and the Appendix Supplements and other suggested readings are at:

http://cda.psych.uiuc.edu/sgep_course_material

the syllabus: [sgep_course_syllabus.pdf](#)

required (longer) readings: [sgep_readings_syllabus_ordered.pdf](#)

appendix supplements and suggested readings:

[sgep_supplements.pdf](#)

this talk: [jsm_ethics_talk.pdf](#)

Also, various slides and handouts used in class will be placed at this site as the course progresses.

The ones that are there now involve:

1) possible class projects for graduate student enrollees:
projects_sgep.pdf

2) synopses of various films to be shown in the class:
films_sgep.pdf

3) sets of “keywords” for the required readings for each week of the course: keywords_sgep.pdf

For Graduate Credit

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

An additional paper is required covering some substantive topic needing original and cogent statistical/probabilistic thought.

For representative topics that might be of interest, see the required and suggested readings in the syllabus (and in `projects_sgep.pdf`);

also, search for relevant statistical articles/applications in the behavioral sciences on the New York Times website: try, for example, Benedict Carey or Gina Kolata.

Graduate students might wish to use this opportunity to generate a publication that might be submitted to an outlet such as *Significance*, the joint (popular) statistics venture between the Royal Statistical Society and the American Statistical Association. See

www.significancemagazine.org

for the type of article they publish (all issues of *Significance* are available on-line through the Illinois library).

Course Topics Sequence (15 Weeks)

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

Week:

- 1 Introduction to (and Review of) the Course
- 2 Probability Theory: Background and Bayes Theorem
- 3 Probability Theory: Application Areas
- 4 Correlation
- 5 Prediction
- 6 The Basic Sampling Model and Associated Topics
- 7 Psychometrics
- 8 Background: Data Presentation and Interpretation
- 9 (Mis)reporting of Data

10 Inferring Causality

11 Simpson's Paradox; Meta-Analysis

12 Statistical Sleuthing and Explanation

13 Background: Experimental Design and the Collection of
Data

14 Ethical Considerations in Data Collection

15 The Federal Rules of Evidence; Some Concluding Remarks

Week 3: Probability Theory: Application Areas

The Ethical Practice of Statistics (for the Perplexed): A Course in the Social and Behavioral Sciences

Lawrence Hubert

— how subjective probabilities might be related to the four levels of a “legal burden of proof”: “preponderance of the evidence”; “clear and convincing evidence”; “clear, unequivocal, and convincing evidence”; and “proof beyond a reasonable doubt”

— the distinction between “general causation” and “specific causation”; the common legal standard for arguing specific causation as an “attributable proportion of risk” of 50% or more

— issues of probability, risk, and gambling; spread betting and point shaving; parimutuel betting; the importance of context and framing in risky choice and decision-making

Required Reading from SGEP (87–118)

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

4.1 Some Probability Considerations in Discrimination and Classification

4.2 Probability and Litigation

4.2.1 Probability of causation

4.2.2 Probability scales and rulers

4.2.3 The cases of Vincent Gigante and Agent Orange

4.3 Betting, Gaming, and Risk

4.3.1 Spread betting

4.3.2 Parimutuel betting

4.3.3 Some psychological considerations in gambling

Popular Articles

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

Better Decisions Through Science, John A. Swets, Robyn M. Dawes, and John Monahan (*Scientific American*, October 2000)

Do Fingerprints Lie? Michael Specter (*New Yorker*, May 27, 2002)

Under Suspicion, Atul Gawande (*New Yorker*, January 8, 2001)

Suggested Reading

The Ethical Practice of Statistics (for the Perplexed): A Course in the Social and Behavioral Sciences

Lawrence Hubert

4.1 Suggested Reading on Agent Orange and Judge Weinstein

4.1.1 Appendix: The Redacted Text of Judge Weinstein's Opinion in the Fatico Case

4.1.2 Appendix: Guidelines for Determining the Probability of Causation and Methods for Radiation Dose Reconstruction Under the Employees Occupational Illness Compensation Program Act of 2000

4.1.3 Appendix: District of Columbia Court of Appeals, In Re As. H (Decided: June 10, 2004)

4.2 Suggested Reading on Issues of Risk

4.3 Suggested Reading on Issues of Betting and Gaming

Film: *The Central Park Five* (2 hours)

An Example of How the Lectures Might Proceed

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

Week 2: Probability Theory: Background and Bayes Theorem

— the case of Sally Clark, wrongly convicted in England of killing her two children; this miscarriage of justice was due to an inappropriate assumption of statistical independence and the commission of the “Prosecutor’s Fallacy”

— breast cancer screening through mammograms; understanding Bayes’ theorem, test sensitivity and specificity, prior probabilities, and the positive predictive value (for example, what is the probability of having breast cancer if the mammogram is “positive”?)

Beginning Quotations

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

Probability theory is nothing but common sense reduced to calculation.

— Pierre-Simon Laplace (*Essai Philosophique Sur Les Probabilités*, 1814)

Misunderstanding of probability may be the greatest of all impediments to scientific literacy.

— Stephen Jay Gould

Probability: Event Independence

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

The formalism of thought offered by probability theory is one of the more useful portions of any beginning course in statistics in helping to promote ethical reasoning.

As typically presented, we speak of an event represented by a capital letter, say A , and the probability of the event as some number in the range from 0 to 1, written as $P(A)$.

The value of 0 is assigned to the “impossible” event that can never occur;

1 is assigned to the “sure” event that will always occur.

The driving condition for the complete edifice of all probability theory is one single postulate:

for two mutually exclusive events, A and B (where mutually exclusivity implies that both events cannot occur at the same time), $P(A \text{ or } B) = P(A) + P(B)$.

As a final beginning definition, we say that two events are independent whenever the probability of the joint event, $P(A \text{ and } B)$, factors as the product of the individual probabilities, $P(A)P(B)$.

The Case of Sally Clark

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

The idea of statistical independence and the factoring of the joint event probability immediately provides a formal tool for understanding a number of historical miscarriages of justice.

In particular, if two events are not independent, then the joint probability cannot be generated by a simple product of the individual probabilities.

A recent example is the case of Sally Clark; she was convicted in England of killing her two children, partially on the basis of an inappropriate assumption of statistical independence.

The purveyor of statistical misinformation in this case was Sir Roy Meadow, famous for Meadow's Law:

“‘One sudden infant death is a tragedy, two is suspicious, and three is murder until proved otherwise’ is a crude aphorism but a sensible working rule for anyone encountering these tragedies.”

We quote part of a news release from the Royal Statistical Society (October 23, 2001):

The Royal Statistical Society today issued a statement, prompted by issues raised by the Sally Clark case, expressing its concern at the misuse of statistics in the courts.

In the recent highly-publicised case of *R v. Sally Clark*, a medical expert witness drew on published studies to obtain a figure for the frequency of sudden infant death syndrome (SIDS, or 'cot death') in families having some of the characteristics of the defendant's family. He went on to square this figure to obtain a value of 1 in 73 million for the frequency of two cases of SIDS in such a family.

This approach is, in general, statistically invalid. It would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically.

More on Sally Clark: The Prosecutor's Fallacy

The Ethical
Practice of
Statistics (for
the
Perplexed): A
Course in the
Social and
Behavioral
Sciences

Lawrence
Hubert

We continue with the Sally Clark episode where the invalidly constructed probability of 1 in 73 million was used to successfully argue for Sally Clark's guilt.

Let A be the event of innocence and B the event of two "cot deaths" within the same family.

The invalid probability of 1 in 73 million was considered to be for $P(B|A)$; a simple equating with $P(A|B)$, the probability of innocence given the two cot deaths, led directly to Sally Clark's conviction.

We continue with the Royal Statistical Society news release:
Aside from its invalidity, figures such as the 1 in 73 million are very easily misinterpreted. Some press reports at the time stated that this was the chance that the deaths of Sally Clark's two children were accidental. This (mis-)interpretation is a serious error of logic known as the Prosecutor's Fallacy.

Sally Clark's conviction was overturned in 2003, and she was released from prison.

Sally Clark died of acute alcohol poisoning in her home four years later in 2007, at the age of 42.

Roy Meadow (1933–) is still an active British pediatrician.

He rose to fame for his 1977 academic article in the *Lancet* on Munchausen Syndrome by Proxy (MSbP); he is the person who coined the name.

He has spent his whole career crusading and testifying against parents, especially mothers, who supposedly wilfully harmed or killed their children.

We quote from Lord Howe, the opposition spokesman for health, speaking in the House of Lords on MSbP (February 2003):

... [O]ne of the most pernicious and ill-founded theories to have gained currency in childcare and social services over the past 10 to 15 years. The theory states that there are parents who induce or fabricate illnesses in their children in order to gain attention for themselves. The name given to it is Münchausen's syndrome by proxy. It is a theory without science. There is no body of peer-reviewed research to underpin MSBP. It rests instead on the assertions of its inventor and on a handful of case histories. When challenged to produce his research papers to justify his original findings, the inventor of MSBP stated, if you please, that he had destroyed them.