ETHICAL STATISTICS AND STATISTICAL ETHICS: THE EXPERIENCE OF CREATING AN INTERDISCIPLINARY MODULE

Lawrence M. Lesser,
Armstrong Atlantic State University,
11935 Abercorn, Savannah, GA 31419

KEY WORDS: Critical Thinking; Ethics; Interdisciplinary; Randomized Response.

Abstract

This article describes an innovative curriculum module the author created on the two-way exchange between ethics and statistics. Part of an undergraduate interdisciplinary ethics course, the module builds upon a prior introduction to basic philosophical ethics, but has no particular mathematical prerequisites beyond high school algebra. Its emphasis on conceptual and critical thinking makes it easily adaptable to high school classrooms as well as readily expandable for more mathematically sophisticated audiences. Through media examples and in-class explorations, the module made connections to contemporary topics such as the death penalty, drug testing, salary equity, and profiling. This article shares resources, strategies and lessons learned for instructors wishing to develop their own modules of various lengths.

1. Introduction

The author created a module that was one-fifth of a 15-week interdisciplinary ethics course undergraduates can take to fulfill a requirement from the "ethics and values" core curriculum category at a comprehensive university in the southern United States. The university catalog describes the course to have "several modules taught by different professors: the first philosophical framework module will provide an explanation and analysis of the principal ethical theories of the western world and subsequent modules will focus on moral issues and case studies in specified areas."

The module has no particular mathematical prerequisites, and so it was not assumed that students had mathematical knowledge beyond high school algebra. In the spirit of authors such as Utts (1996), the module emphasized conceptual and critical thinking more than formulas. The most complex mathematical skill absolutely required was setting up and (with the help of a calculator) evaluating a proportion raised to an integer power. This module is therefore easily adaptable to high school (and even some middle school) classrooms as well as readily mathematically enriched for upper division undergraduates or majors in the mathematical sciences.

This aspect of the module was designed to reinforce the common initial 3-week module of the course, which laid the philosophical foundation for two main examples of normative ethics: Kantian and utilitarian. Although there certainly are primary source references available for these (e.g., Kant (1959) and Mill (1979)), students in this course were given a condensed introduction in the opening common module as well as access to more detailed notes on reserve in the library.

The author's module was delivered primarily with a discussion and activity-based format, appropriately reinforced with lectures and case studies, and students were required to bring a calculator to each meeting. The topics began by roughly following the sequence of steps in a quantitative study, exploring ethical issues in collecting data from human subjects, continuing with the ethics of displaying and reporting findings, and then making applications to social issues in the bigger picture. After five regular meetings (paralleling this article's sequence of sections 3.1 through 3.5, the sixth meeting was a cumulative (multiple-choice and essay) exam. The author taught the course mostly from personal notes, handouts and online resources to supplement the required textbook Huff (1993).

2. From Ethical Statistics to Statistical Ethics

The author had the students access and discuss the guidelines the ASA developed in 1983, adopted in 1989, and made available on the World Wide Web at: www.amstat.org/profession/ethicalstatistics.html. The more unique element of the module, however, was to show not just the professional ethics of the statistics profession, but also how statistics could be applied as a tool to engage in some meta-ethics (analyzing the meaning of ethical claims), and thus help people face ethical and social issues affecting all of society. Statistics clearly has a role to play to understand how things are constituted -- what data reveals about hiring, salaries, racial profiling, death penalty, etc. Particularly in an ethics module, class discussion can include both calculations and implications. As the 19th-century French statistician Eugene Buret said, "It no longer suffices to know how things are constituted: we need to seek how things should be constituted so that this world of ours may present less suffering and destitution."

A goal of the module is to apply statistics to real-world scenarios to show both (1) how unavoidably useful quantitative approaches can be in exploring them and yet (2) the limitations of precise implementations or interpretations of common ethical principles.

3. Outline of Module Topics

3.1 Ethics in Collecting Quantitative Data from Humans

Various ways were discussed how the students might randomly assign one of two treatments to each person in the room, followed by discussing when random assignment of treatments would be ethical (or even possible). It was also discussed when, if ever, it would be ethical for a person to be in a study without his consent and/or knowledge (either of the true purpose of the study or that there is a study at all!). Students accessed online descriptions from the United States Holocaust Memorial Museum about the experiments conducted by German physicians on thousands of concentration camp prisoners during World War II. Students then appreciate the need for the Nuremburg Code, and the motivation for some of the language in their university's IRB guidelines or in the ASA ethics codes. Most students are shocked to learn, however, of examples of unethical mainstream medical research that occurred after the Nuremberg Code, such as the decades-long Tuskegee Syphilis Study (Jones 1993).

By now students accept the idea that some treatments or traits are not appropriate (or even possible) to assign randomly to subjects and naturally have prepared themselves for the concept of studies in which pre-existing characteristics are observed. Rich classroom discussions are held on "obvious" examples that should be done only as observational studies (e.g., smoking while pregnant and birthweight of resulting baby) and others that are not quite as obvious (use of cell phones and brain cancer). The class then discovers that even the survey method has room to further protect the privacy and dignity of individual participants while obtaining accurate collective information on sensitive or personal behaviors. While time constraints limit the interactive demonstration of randomized response (e.g., Warner 1965; Burger and Starbird 2000; Bolstad, Hunt and McWhirter 2001) to its simplest form, this example has immediate impact and interest for the module students.

3.2 Ethics of Displaying and Reporting Findings

Having addressed issues of data obtained ethically, students must now be on guard for unethical ways data may be presented. Huff (1993) gives examples of gee-whiz line or bar graphs with an incomplete vertical axis, as well as the pitfall of pictographs whose areas are not proportional to their heights (and thus, to the numbers they supposedly represent). Additional examples from current real-life mass media are also explored.

Pitfalls of numerical summaries are then discussed, such as the many types of average. As an in-class activity, students are given a small number of "class sizes" and simply asked to determine that university's "average class size". Students did not realize that even choosing the arithmetic mean as the type of average allows very different results if one averages per student instead of per class (e.g., Hemenway 1982). One student suggested discarding the "outlier" class size, triggering an unplanned insightful discussion of when one can justify discarding outliers.

To show an example of deceptive (and therefore unethical, at least to a Kantian) interpretation of a probability, I ask for a $1 bill from a student, read off the numerical part of the serial number, then dramatically proclaim the occurrence of an event with a 1 in 100 million probability! While students were not "fooled" by this, they did not readily articulate a formal explanation or generate examples of how the simple issue of whether events are fully specified in advance has been raised in the context of more subtle and weighty contexts (e.g., Simon 1998). The more familiar pitfall of equating causation and correlation was also explored.

3.3 Applying Statistics to Ethical/Social Issues

Ethical issues in hiring are under great scrutiny these days. Certain forms of affirmative action may be viewed to benefit society overall (and thus be supported by utilitarians, especially "rule-utilitarians") even though they may treat some individuals unfairly in a way a Kantian would oppose in the process. The need to assess whether society is better off motivates exploring the power and limitations of analyzing hiring statistics. We can help students experience that a single number is rarely a sufficient picture of a data set by, for example, exploring a dataset in which males are favored at the aggregate level even though females are favored at all sublevels (Lesser 2001). A Kantian would certainly be more likely than a utilitarian to insist that a charge of discrimination be based on intent rather than on numerical hiring results alone. Even utilitarians, however, have a nontrivial task to decide what numerical basis to use.

Fairness in salaries has also received much attention recently. "Equal pay for equal work" was a phrase often heard during the Presidential 2000 election campaigns

and debates. To arm students with the kind of information statistical figures could reveal or possibly conceal regarding salaries, we ask students to consider a company which employs two categories of workers -- executives and support staff (students could also examine this dataset as if it represented two types of workers, regardless of company, such as educators and lawyers!), and readily verify that females are paid better within each of the two categories, but less overall. This type of example allows students to understand the statement of Vos Savant (2000, p. 14), who reports that the commonly quoted Bureau of Labor Statistics data that women earn about 77% of what men earn "simply compares the weekly median earnings of all working women and all working men....It has nothing whatsoever to do with equal pay for equal work. Instead, it merely indicates that men generally occupy positions that pay more."

The class then explores a dataset in which all males and females are the same type of worker, and whose salaries can be explained perfectly by their varying years of experience. Finally, to explore what can happen when raises are tied purely to performance, the class explores the implications of giving teachers raises based on how well their students test at the end of the year. Students recognize that the class with the higher end-of-year score may not have had the highest improvement or have had the same theoretical room for improvement. Furthermore, some teachers may coach unduly or "teach to the test", the ethics of which have been discussed (e.g., Meike 1996, George 1987).

3.4 Applying Probability to Ethical/Social Issues

This is one of the few parts of the module in which probability laws (complement rule and the multiplication rule for independent events) are invoked, but in a manner not needing formal symbolism (e.g., Utts 1996). The death penalty is a topic that has received a new wave of attention, thanks to the recent Presidential debates as well as the delayed execution of Timothy McVeigh. Ambiguities due to aggregation can be revisited here (e.g., Moore and McCabe 1999, p. 207). Some defenders (e.g., Pambianco 2000) of the death penalty try to strengthen their position by claiming that no innocent person has been executed. We then conduct a straightforward calculation that shows the extreme implausibility of this claim, using 4458 as the number of US prisoners executed under civil authority during 1930-1999 (source: 2000 *Statistical Abstract of the US*) and the highest of the student-suggested probabilities that a particular trial was a correct decision. There are types of arguments (especially utilitarian ones) for the death penalty, of course, that might be unaffected by this result.

The statistics of drug testing are explored, giving students context to discuss the statement by Paulos (1988, pp. 66-67): "To subject people who test positive to stigmas, especially when most of them may be false positives, is counterproductive and wrong."

Yet another topic that has been in the 2000 Presidential debates and other recent news (e.g., McAplin 2000, Kinsley 2001) is profiling. Because statistics is designed to gain information about groups more than individuals, the issue of when it may be unethical to treat an individual as a member of group is of particular interest. Some, but not all, students distinguished whether race is a component in a profile from whether it is the only component. From some points of view, profiling is neutral or perhaps a necessary evil, from other viewpoints, simply evil. A Kantian would consider profiling immoral to the extent that it denies a person's personhood or reduces a person to an object. The students explore what statistics might be involved in determining the presence or nature of profiling, such as how the percentages that are minorities of those stopped, searched and arrested compares to the percentage of total miles driven by minorities or how it compares to the percentage of major accidents caused by minorities. Students also compare and contrast this example of profiling with instances of profiling practiced by auto insurance companies, university admissions officers, and airport security officials.

3.5 Connections to Decision Making

While the two previous parts of the module applied statistics to real-world scenarios of ethical importance, the focus of this part is the explore the role statistics can play in philosophical ethics itself, completing the connection to the opening common module of the course. One key example of this analyzed the practical issues in implementing the utilitarian goal of "greatest happiness for the greatest number."

Adopting a familiar introductory philosophy format of "which of these worlds is better?", students were presented with the "amount of happiness" each of five individuals would have in each of three hypothetical worlds. The author chose the amounts so that the mean was identical in all three worlds, but the world in which the greatest absolute good was achieved (by someone) was not the world in which the greatest local good was achieved by the greatest number of people, etc. Not only was there a way to justify a choice of any of the three worlds based on the interpretation and weights you assigned to the two instances of "greatest", but also no

matter what world you picked, there was always another world that would make a majority of the five people better off!

4. Lessons Learned

In end-of-course written evaluations, many students expressed satisfaction at how relevant the module was to everyday life (including one student who admitted "When I saw it on the syllabus I thought it was going to be boring."), as supported by the many recent media clippings they encountered in the class. Students seemed surprised by just how much overlap there was between the realms of ethics and statistics. Some overlap was immediate while other aspects required more sustained reflection on the question: "What would (Kant / Mill) say?"

On the one hand, tools such as randomized response gave students new respect for the power of statistics to estimate quantities that cannot be accurately or readily observed. On the other hand, students had new appreciation for the types of ways in which statistics could be misused or abused. They gained appreciation for the balance recently articulated by Best (2001, B8): "Some statistics are bad, but others are pretty good, and we need statistics -- good statistics-- to talk sensibly about social problems. The solution, then, is not to give up on statistics, but to become better judges of the numbers we encounter." Movement towards such a view was evidenced by the end of the course, when students reached a strong consensus that statistical evidence alone should rarely suffice to convict, but should justify launching an investigation.

Students who may have been quick to give numbers an unquestioned authority to describe or prescribe social issues were forced to confront their tendency by seeing several examples in which the conclusion varied with the angle of analysis (e.g., type of average, level of aggregation). We found, as did Oberhoff and Barnes (2000, p. 50), that our examples "generated considerable student discussion, and it is important to point out that, in general, there does not always exist a 'best solution' to an ethical dilemma." Younger college students may not be used to encountering ambiguity, especially in mathematics or statistics classes, and some would often ask insistently "So for the scenario you showed us, what is the answer?" The instructor found it important to encourage students to make explicit what their own assumptions and values are in a way that distinguished between the consequences of their statistical analysis and those of their ethical analysis.

The instructor made a conscious choice to facilitate discussion more than simply lecture, in support of Bok (1976), who notes that students in lecture-dominated ethics classes do not develop the power of moral reasoning. Bok (1976, p. 30) warns, however, that the instructor must "know how to conduct a rigorous class discussion that will elicit a full consideration of the issues without degenerating into a windy exchange of student opinion." The author's attempt to create an interactive environment appeared validated by end-of-course student evaluations (n = 81 responding) in which 69% strongly agreed (and 97% agreed or strongly agreed) that "the instructor encouraged students to express themselves freely and openly and to question and discuss the issues presented in class." Open-ended written evaluation feedback consistently described the class as "very enjoyable", related to everyday life, and revealing pitfalls previously unknown to the student.

As mentioned earlier, a major challenge was the lack of a textbook that covered the application of statistics and ethics to one another. Bok (1976) discusses the importance and the difficulty of an instructor having sufficient background in pedagogy, moral philosophy, and the area to which ethics will be applied. While the instructor felt comfortable (as did Webber 1997) after doing some reading on his own and using a member of the philosophy department as a resource, there were students (especially those whose attendance is inconsistent) who were not used to have so much of a course be "beyond the book" interactive discussion. The instructor deliberately did not try to pack quite as much required material into the course as he would have if students had more self-contained resources to fall back on to supplement class meetings.
find statistics to be one of the most rich and universal vehicles to implement an interdisciplinary ethics requirement. Statistics or mathematics departments desiring a required ethics experience specific to their majors may likely share the situation of Oberhoff and Barnes (2000), who report that the large number of required courses in their degree programs did not allow offering an entire course in ethics. We hope we have offered an outline and sampling of meaningful connections between statistics and ethics that could be used as a basis for a module of any length, up to a full course.

5. References

Best, J. (2001). Telling the Truth About Damned Lies and Statistics. *The Chronicle of Higher Education*, May 4, 2001, pp. B7-B9.

Bok, D. (October 1976). Can Ethics Be Taught? *Change*, pp. 26-30.

Bolstad, W.M.; Hunt, L.A., McWhiter, J.L. (May 2001). Sex, Drugs and Rock & Roll Survey in a First-Year Service Course in Statistics. *American Statistician*, 55, 2, 145-149.

Burger, E. B. & Starbird, M. (2000). *The Heart of Mathematics: an Invitation to Effective Thinking*. Emeryville, CA: Key College Publishing.

George, P. (Sept. 1987). Soundoff-- Coaching for Standardized Tests: Efficacy and Ethics. *Mathematics Teacher*, 80 (6), 424-426.

Hemenway, D. (1982). Why Your Classes are Larger than 'Average.' *Mathematics Magazine*, 55(3), 162-164.

Huff, D. (1993). *How to Lie with Statistics*. New York: W.W. Norton.

Jones, J. (1993). *Bad Blood: The Tuskegee Syphilis Experiment*. New York: Free Press.

Kant, I. (1959). *Foundations of the Metaphysics of Morals*, [translation from the original in 1785]. Indianapolis, IN: Bobbs-Merrill.

Kinsley, M. (October 2, 2001). Some forms of rational discrimination may be acceptable to ensure safety, *Savannah Morning New*s, p. 9A.

Lesser, L. (2001). Representations of Reversal: Exploring Simpson's Paradox. In A. Cuoco and F. Curcio (Eds.). *The Roles of Representation in School Mathematics*, pp. 129-145 Reston, VA: National Council of Teachers of Mathematics.

McAplin, J. (October 13, 2000). Memos: N.J. Knew of Racial Profiling, *Savannah Morning News*, p. 9A.

Meike, G. (1996). Soundoff: Coaching Versus Competence. *Mathematics Teacher*, 89 (4), 270-272.

Mill, J.S. (1979). *Utilitarianism* [edited version of original 1861 work]. Indianapolis: Hackett Publishing.

Moore, D. and G. McCabe. (1999). *Introduction to the Practice of Statistics* (3rd ed.). New York: W. H. Freeman.

Oberhoff, K. and R. Barnes. (Sept. 2000). Senior Seminar: A Capstone Course in the Computer and Mathematical Sciences. *Humanistic Mathematics Network Journal*, 23, 49-54.

Pambianco, R. (June 23, 2000). Innocents Are Being Executed? Name One. *Savannah Morning News*, p. 17A.

Paulos, J. (1988). *Innumeracy: Mathematical Illiteracy and its Consequences*. New York: Hill and Wang.

Simon, B. (1998) Torah Codes. *Jewish Action*, Vol. 58, No. 3.

Utts, J. (1996). *Seeing Through Statistics*. Pacific Grove, CA: Brooks/Cole.

Vos Savant, M. (Dec. 31, 2000). Ask Marilyn. *Parade Magazine*.

Warner, S. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, March 1965, Vol. 60, No. 309, 63-69.

Webber, R. (1997). A Course in Mathematical Ethics. *Humanistic Mathematics Network Journal*, 16, 48-49.
Acknowledgments:

This work was supported in part by the author's Arthur M. Gignilliat, Jr. Professorship at Armstrong Atlantic State University and was presented (in part) at the 2001 Joint Statistical Meetings. The author also expresses appreciation to Dr. Bonnie Shulman and Dr. Erik Nordenhaug, and also to J.B. Lesser for assistance with transportation.