Ethical Statistics and Statistical Ethics: Making an Interdisciplinary Module

Lawrence M. Lesser
University of Texas at El Paso

Erik Nordenhaug
Armstrong Atlantic State University

Journal of Statistics Education Volume 12, Number 3 (2004), www2.amstat.org/publications/jse/v12n3/lesser.html

Copyright © 2004 by Lawrence M. Lesser and Erik Nordenhaug, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Critical thinking; Curriculum; Kantian; Philosophy; Statistical reasoning; Utilitarian

Abstract

This article describes an innovative curriculum module the first author created on the two-way exchange between statistics and applied ethics. The module, having no particular mathematical prerequisites beyond high school algebra, is part of an undergraduate interdisciplinary ethics course which begins with a 3-week introduction to basic applied ethics taught by a philosophy professor (the second author), and continues with 3-week modules from professors in various other disciplines. The first author’s module’s emphasis on conceptual and critical thinking makes it easily adaptable to service-level courses as well as readily expandable for more mathematically sophisticated audiences. Through in-class explorations and discussions, the module made connections to contemporary topics such as the death penalty, equal pay for equal work, and profiling. This article shares examples, resources, strategies and lessons learned for instructors wishing to develop their own modules of various lengths.

1. Overview of the Course and Module

The author created a module that was one of five modules comprising a 15-week interdisciplinary ethics course undergraduates can take to fulfill a requirement from the “ethics and values” core curriculum category at Armstrong Atlantic State University, a comprehensive masters university of 6,000 students in the southern United States. The overall ETHC 2000 course syllabus lists three objectives: “To introduce students to the basic types and structures of ethical reasoning, to apply ethical reasoning to various professions and areas of academic training, and to challenge the student to think deeply about ethical issues as they arise in society.” In this article, the word “module” (not “course”) refers to the part of the multi-instructor, multi-module course taught by the first author, and the words “author” and “instructor” refer only to the first author unless noted otherwise.
The first three-week module of the interdisciplinary course is taught by a professor of philosophy and introduces the basic philosophical framework of the principal ethical theories of the western world, providing familiarity with basic ethical terms such as ethical relativism, cultural relativism, ethical absolutism, egoism, altruism, utilitarianism, deontology, and virtue.

After this large-group opening module, the students (about 150 each semester) are then broken into four smaller classroom sections (of roughly 35 students each) that rotate through a round-robin of four 3-week modules taught by different professors in a variety of disciplines. Recent module titles include: bioethics, Confucian ethics, information age ethics, quality of life issues in health care, and ethical issues of death in Melville and Tolstoy. Each student therefore experiences a total of five modules (unless she exercises an option to omit one of the last four modules for one less credit hour). Each professor (except the philosophy professor) delivers his module four consecutive times during the semester.

Each individual module professor had fair latitude to decide her readings and attendance/assessment policies. Because of the short time frame, a three-week module typically had one main formal assessment (a paper or an exam) on or due by its sixth (and final) meeting, though homework (e.g., readings) were assigned each class meeting and attendance directly impacted the module grade as well. A student’s course grade for the overall interdisciplinary course was then determined by the unweighted arithmetic mean of the scores for all five modules.

With his particular module, the author certainly hoped to give students the opportunity to explore the ethical pitfalls of misapplied statistics. He also wanted to explore the reverse direction of applied analysis between the two fields (ethics and statistics), not unlike Rose (1996, p.15): “Alternatively, perhaps complementarily, instructors would provide students with examples of situations in which statistical tools enable them to evaluate particular ethical questions, such as discrimination or a ‘just’ distribution of resources. Such examples could lead to a discussion on the extent to which ‘fairness’ might be quantified or how quantitative analysis can help to evaluate fairness.”

The module (and overall course, for that matter) has no particular mathematical or statistical 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 a whole number power, and the author has since successfully taught most of the mathematical content of this module to high school students in Texas. (It should be noted that having the formal mathematical prerequisite knowledge does not ensure the conceptual maturity needed to handle all of the critical thinking.) The instructor also gave well received presentations on this material for the Armstrong Atlantic State University’s Philosophical Debate Group, none of whose members had taken or were taking the module. As will become clear, there are many places where examples can be explored in greater depth or mathematical detail to flesh out a university course more suitable for courses or majors in a mathematical sciences department.

The module was delivered with a discussion-based format, appropriately integrating mini-lectures and case studies, and students were required to bring a calculator to each meeting to verify or explore the quantitative aspects of certain scenarios. Much of the class time roughly paralleled procedures identified in moral education (e.g., Arbuthnot and Faust 1981), by presenting an ambiguous scenario for a topic, obtaining initial opinions, guiding a discussion or debate to draw out as many contrasts and distinctions as possible before articulating a resolution based on a specific ethical theory. The topics began with exploring ethical issues in collecting data, then continuing with the ethics of displaying and reporting findings (just as data collection precedes data analysis and reporting in a typical study), and then making applications to social issues in the bigger picture. After five regular meetings (whose topics mirror this article’s sequence of sections 3.1 through 3.5), the sixth meeting was a cumulative exam with both multiple-choice and essay questions. Some questions from this exam are shared throughout this article.
As Webber (1997) notes, there do not seem to be books available on mathematical and statistical ethics as there are for ethics in computer science, for example. The instructor therefore taught the course mostly from personal notes, handouts and online resources to supplement the required textbook Huff (1993), which has been in print since 1954, attesting to its status as a “classic” for broad audiences. This 142-page paperback was assigned (with selected highlights briefly discussed in class) for its colorful illustrations of statistical misuse and abuse, but even more to give students a user-friendly introduction and context for some key statistical terms and concepts that would be referenced in discussion from more pertinent or developed scenarios encountered later in the module. For example, Huff’s first chapter familiarizes students with some basic terms and vocabulary about sampling and then poses a question that launches the module’s randomized response discussion (see Section 3.1.3).

2. Background on Ethics

2.1 Professional Ethics for Statistics

Having worked as a statistician inside and outside academia, the author found the professional ethics part a more familiar domain than philosophical ethics itself. Beginning with the August-September 1993 issue, there have been many columns of “The Ethical Statistician” in AMSTAT News that present case studies and raise general and specific issues related to ethics that can be worked into class discussions. Continuing Education classes in Ethical Statistics have also been held at recent Joint Statistical Meetings, and “Data Ethics” chapters are appearing in introductory books (e.g., Moore 2001; Utts 2005). An example of general background is a column by Griffith (1995) that distinguishes between “ethics”, “morals” and “professional competency.” The instructor asked the students to access and review the ASA’s 1999 “Ethical Guidelines for Statistical Practice” (and one module exam essay question asked students to name an example of a significant issue from it). A Related Websites section at the end of this article lists the URL for the ASA Guidelines and for other examples of published codes, including the Nuremburg Code, the Belmont Report, the Declaration of Helsinki, the International Statistical Institute’s “Declaration on Professional Ethics” and the United Nations statistical division’s “Fundamental Principles of Official Statistics”.

By spending some time examining such codes, students will appreciate that there are many important codes that help guide statistical work and statisticians, covering a broader range of issues than students typically anticipated. It is also worth having students note that we cannot say there is complete agreement as to all elements or emphasis. Indeed, in Section A of the Preamble of the ASA (1999) guidelines, there is the statement: “The guidelines may be partially conflicting in specific cases.” A further caution about any attempt to reduce ethics to a set of simple rules is that it may place more emphasis on the letter of the rules than their underlying ethical principles (Seltzer 2001).

Because students taking this module (and course) were not required to have had a prior course in statistics, some of the initial examples were simply familiar examples (e.g., from Huff 1993) of (often intentional) statistical misuse. Such examples may not appear to have a highly sophisticated ethical dimension in that there usually seems to be little room for ethical controversy when data are presented and interpreted in accordance with conventional guidelines of professional practice. The discussion of these examples is designed to give students (or remind, for those who have been exposed in college or high school to) just enough concepts and vocabulary that they can go beyond simple examples of misapplied statistics and discuss more substantial and/or controversial scenarios involving statistics and ethics. In his ASA President's Invited Address at the 2001 Joint Statistical Meetings, Dr. Robert L. Frosch (2001, p. 16) asserts that “the statistician has to be an ethicist, and must think about pushing problems and solutions at people so that they understand what is involved, and not just react to what others think is involved....You have to follow the problems where they lead.” Some background on applied ethics is provided in the next section for readers less familiar with this topic. A possible exam essay question is to discuss the claim by
Vardeman and Morris (2003, p.21): “[Statistics’] real contribution to society is primarily moral, not technical.”

To make this theme more concrete, students may find it interesting to hear or read about specific individuals such as Roger Boisjoly, who received the Prize for Scientific Freedom and Responsibility from the American Association for the Advancement of Science for “for his honesty and integrity leading up to and directly following the [1986 Challenger] shuttle disaster” (see Related Websites).

As this article now transitions from professional ethics for statisticians to examining ethics for all people, it is valuable to reflect upon the words of Sia (2001, pp. 69-70): "... while certain ethical considerations do arise in specific professional contexts, e.g., engineering, medical, business, the primary context is our human nature. That is to say, ethical questions arise because we are first and foremost human beings in search of answers as to how we ought to behave..." Sia argues that it follows that "... the teaching of ethics by its nature not only cuts across the whole curriculum but also undergirds it."

2.2 Major Schools of Applied Ethics in Philosophy

While there is overlap between ethics and professional or legal codes, it is important to distinguish them, since codes are supposed to be based on moral standards, not vice versa. A distinction between right and wrong in ethics is based on reason and logic, while a code is ultimately what a large group of people state and agree to follow. The importance of having one’s ethics knowledge include more than the nuts and bolts of professional ethics is underscored in the advice to statisticians from Gardenier (2002, p. 22): “You can make your case persuasively if you know not only what the statistical ethics guidelines are, but also how they relate to professional ethics generally and their basis in ethical philosophy.” Familiarity with the major schools of thought in the theory of normative (as opposed to descriptive) ethics must, at a minimum, include the utilitarian and Kantian approaches.

Briefly, the family of ethical theories referred to as utilitarianism generally judges the goodness of actions by their consequences (i.e., by how much utility or happiness is produced), giving all individuals equal consideration. Act-utilitarians choose the particular act (while rule-utilitarians act according to the rule) which would result in the “greatest good for the greatest number.” One aspect of act-utilitarianism is captured in the witty words of a t-shirt sold by the American Statistical Association: “The N's justify the means.”

Kantianism, however, always treats other persons as ends, not means, and judges the rightness of actions not by consequences but by whether the person followed the (unconditional) categorical imperative "Act only according to that maxim whereby you can at the same time will that it become a universal law." Thus, Kant focuses on motive (deontological ethics, not cosmological) and rationality as the basis of moral obligation.

Readers of this article desiring more background on these two schools may consult the Appendices (written by the second author, a philosophy professor with particular background in ethics) or primary source references (e.g., Kant 1959; Mill 1979). Students in this course were given a condensed introduction in class as well as access to more detailed notes on reserve in the library. Students are encouraged to reflect on how neither school may seem satisfying all the time in all cases, but each school offers extremely valuable insights in its application, has made huge contributions to society, and the two schools together form conceptually useful juxtapositions (e.g., intentions versus consequences) for philosophical points of reference and comparison. Exploring Kantian and utilitarian ethics in context may offer a framework to help students identify and classify philosophical approaches and be more likely to recognize when they are not being applied in a consistent manner. Students often see rule-utilitarianism as embodying some of the best qualities of both Kantian and act-utilitarianism, and thus offering reconciliation to what may appear as a sometimes exaggerated juxtaposition.
Of course, there are other schools of ethics (e.g., Aristotle’s virtue ethics, described in Appendix C) that would complement Kantian and utilitarian ethics, but it was not possible in the large-lecture opening module (taught by the philosophy professor who is the article’s second author) of the course to cover more than those two schools with reasonable thoroughness in only 3 weeks (i.e., 5 non-exam meetings). (Because of this limitation of this time frame, it was recently decided to restructure future iterations of the class so that there will be roughly four 4-week modules instead of five 3-week modules, and this should also help address other instances in this article that refer to time constraints.) Also, AASU offers a full semester introduction to philosophical ethics course for students wanting to follow up with a more detailed and comprehensive exploration.

2.3 How Utilitarian and Kantian Ethics Connect with Statistics

As mentioned in Section 2.1, the more unique element of this module was to go beyond the straightforward examples of “misapplied statistics” that might be encountered in any introductory statistics course. This higher goal included gaining some familiarity with some ethical considerations in the practice of statistics and with 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. Our attitude was that asking “What does statistics say?” is enhanced by also asking questions such as “What would Kant say? What would Mill say?” (One exam essay question asked: “If Kant taught, would he be more likely to use norm-referenced assessment or criterion-referenced assessment? State and defend your answer.”)

Having an accurate way to describe (or predict) data is important in sizing up the state of things with respect to important social issues with underlying ethical content. Students are given a selective (mostly conceptual) overview of how statistics are useful in identifying numerical discrepancies, whether in salaries or in executions. Although this sampling is limited severely by the time constraints of this particular module, an instructor with more time (or with students of stronger background) could certainly add more depth, detail, and formality (e.g., create a benchmark by subjecting the comparisons of means or proportions to formal tests of hypotheses).

3. Module Outline of Topics

3.1 Ethics of Collecting Data from Human Subjects

3.1.1 Experiments

The most convincing statistical evidence of causality is generally acknowledged to come from tightly-designed experiments, such as a double-blind, placebo-controlled, randomized clinical trial carried out on a large enough number of people for statistical significance. Such trials are among the most important applications of statistical methods and have yielded the field of medicine many of its most important pieces of knowledge.

Students must first recognize that there is potential tension in the “differences in purpose between clinical trials, which are devoted to generating scientific knowledge with the aim of improving treatment of future patients, and medical care, which is devoted to the diagnosis and treatment of particular patients...Clarity [about these differences] is ethically important at all stages of clinical research, from study design, to subject recruitment and selection, to informed consent, to monitoring the condition of research participants, to deciding when to stop research participation because of adverse events or clinical deterioration.” (Miller 2002, p. 1821). Harvard Medical School’s Dr. Charles Hennekens articulates it: “On the one hand, there must be sufficient belief in the agent’s potential to justify exposing half the subjects to it. On the other hand, there must be sufficient doubt about its efficacy to justify withholding it.
from the other half of subjects who might be assigned to placebos (Moore 2001, p.114).” Because prioritizing the greater good over the welfare of a trial’s individual subjects is clearly a utilitarian idea, utilitarianism has been described as the ethical foundation or justification for randomized clinical trials.

To work up to the above discussion, the class first reviewed much less subtle examples that suggested to students (in the spirit of rule-utilitarianism if not Kantianism) that there are certain experiments that should never be run on an individual, no matter how big the benefit for how many other people. After reviewing some preliminary background and terminology (e.g., random assignment, confidentiality versus anonymity, consent), the class discussed experiments in not-so-distant history that clearly violated Kantian ideals of human dignity. Students accessed online descriptions (see Related Websites) 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 had greater appreciation of the need for the Nuremburg Code, and the motivation for some of the language in their university’s Institutional Review Board guidelines or in the ASA ethics codes, such as the requirement for voluntary consent, the right of the subject to quit the experiment, the avoidance of unnecessary suffering, etc. A portion of the Nuremburg Code that generated particularly lively discussion was: “No experiment should be conducted where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.”

A further ethical angle for discussing the Nazi experiments is whether to use that data. Moe (1984) indicates that at least 45 research articles published since WWII cite data from the Nazi experiments. This has been controversial, as some scholars consider the data morally tainted and possibly scientifically flawed as well, while others argue for an overriding interest in using any information that could save future lives, especially if accompanied by a clear condemnation of the methods.

Most students were shocked, however, to learn there has been unethical mainstream medical research since the Nuremberg Code, most notably the Tuskegee Syphilis Study (Heller 1972; Jones 1993). From 1932-1972, about 400 (mostly poor and illiterate) African-American male sharecroppers in Macon County, AL were told by US Public Health Service physicians that they were being treated for “bad blood” rather than that they had syphilis. Treatments (e.g., penicillin, antibiotics) that became available were withheld (often replaced by “sham treatments” such as aspirin) so that doctors could continue observing the progression of untreated syphilis to see if it was “different” for African-Americans than for Caucasians. Despite the suffering and deaths from this experiment, it seems little new useful knowledge was gained about untreated syphilis, given the existence of a cure (penicillin), making the experiment unethical from utilitarian grounds as well as from Kantian grounds.

It should not be assumed, however, that a Kantian would automatically rule out all experiments. Thiroux (2001, p. 388) explains that a Kantian can preserve the autonomy and inherent value of the human as an end in herself by arguing that: (a) “A person generally should not be experimented on unless the experimental procedure is therapeutic and not harmful” and (b) “A person may be experimented on provided he or she is fully informed and can freely consent to such experimentation, realizing that it may not be therapeutic but good for humanity.”

Also, through discussion of specific examples of valid research, students come to understand that it is neither necessary nor productive to ban all experiments that involve deception or incomplete disclosure, giving further appreciation for ASA Ethics Guidelines such as II.D.8 (of ASA 1999; see Related Websites): “Where full disclosure of study parameters to subjects or to other investigators is not advisable, as in some randomized clinical trials, generally inform them of the nature of the information withheld and the reason for withholding it. As with deception, assure independent ethical review of the protocol and continued monitoring of the research.”
Students typically would think any particular experiment could be classified as simply either ethical or not ethical, and it was a new experience for them to consider how it might be ethical to begin a particular study, but yet not to continue that study beyond a certain point. Students with highly advanced backgrounds might wish to explore some of the complex ethical and statistical details that help guide decisions as to when an ongoing clinical trial should be stopped for efficacy or safety reasons.

As time permits, students can further explore ethics of experiments in other documents (see Related Websites) such as the 1979 Belmont Report and the 1964 Declaration of Helsinki, most recently amended by the World Medical Association in 2000. One statement of the latter suggests increased resistance to utilitarianism: “...considerations related to the well-being of the human subject should take precedence over the interests of science and society.” A more radical specific principle of the latter code questions the use of testing a new method against a placebo (no treatment) when an established method exists that would be superior to a placebo. Because of time constraints, the module did not address ethical dimensions of experiments with animals.

### 3.1.2 Observational Studies

By now, students accept the idea that some treatments or traits are not appropriate (or even possible) to assign randomly to subjects and are now prepared for the concept of studies in which a preexisting risk factor (e.g., whether mother smoked while pregnant) is observed along with an outcome variable such as baby’s birth weight. Rich classroom discussions are held on obvious examples that should be done only as observational studies (e.g., how baby’s birth weight varies with whether the mother smoked while pregnant) and others that are not as obvious (e.g., rate of brain cancer by whether or not cell phones were used regularly). In addition to the situations in which ethical considerations compel an observational study to be done instead of an experiment in the first place, researchers performing observational studies have the general duty to guard against all forms of bias or confounding, and making sure that any conclusions are appropriately qualified.

The instructor needs to make sure students recognize that establishing causation via observational study is much more difficult than via experiment and requires more than mere statistical association, including: having a reasonable model to explain cause and effect, showing that the relationship occurs under various conditions, and ruling out potential confounding variables. [More detail on causation conditions can be found in papers such as Bradford-Hill (1965) and Cornfield's necessary conditions for confounding factors are discussed in Schield (1999).] Students need to appreciate that the near impossibility of being sure one has thought of (and collected data for) all possible factors in the first place is why we prefer when possible to use randomization, which (with more and more likelihood with larger trial sizes) creates groups that on average are similar in all variables -- known and unknown.

Indeed, only in exceptional cases has data from studies other than experiments yielded a broad agreement about causation. We discussed this not just in abstract generality, but in the context of the famous controversy about whether smoking causes lung cancer. First, students discuss the pitfalls of conclusions about human lung cancer based on experiments in which tobacco products produced skin cancer in various animal species, such as the mouse-painting studies of Wynder, Graham, and Croninger (1953, 1955).

Next, we can look at data Cook (1980) supplies from the Doll and Hill (1950, 1952, 1956) retrospective case-control study which looked at smoking behavior among lung cancer patients at a London hospital and a group of hospital patients without lung cancer but with similar distribution of gender, age, etc. When a journal editorial made too strong a claim of causation, noted statistician Ronald Fisher vigorously insisted more analysis was necessary to explain the risk from inhalation and to rule out alternative explanations such as 'cancer causing smoking' or 'individual genetic traits causing both cancer and smoking'.

Excluding alternative explanations is indeed a huge challenge, and students might be given a chance to
discuss how they might do this before the instructor discusses with them some of the actual followup studies done (e.g., studying monozygotic (identical) twins where only one smoked). Another major theme is establishing coherence for the “smoking hypothesis” -- e.g., people decrease their risk by smoking less, smoking for fewer years, and by quitting. (Further plausibility came later with the benefit of additional decades of smoking and cancer data, in which rises in cancer rates happened roughly 30 years after rises in smoking rates.) A concise overview of this appears in Freedman (1999).

A noted statistician who addressed weaknesses in observational studies was William G. Cochran, who served on the Surgeon General’s committee that analyzed seven large prospective studies started in the 1960’s and that wrote a landmark report (Advisory Committee to the Surgeon General of the Public Health Service 1964) that identified smoking as a cause of lung cancer. Students can gain general appreciation about the usage and limitations of observational studies in programs 11 and 12 from Annenberg/CPB (1989), shown in class if time permits or assigned as homework to watch in the library.

3.1.3 Survey Samples

After first reviewing basic terminology and concepts of sampling encountered in the first chapter of Huff (1993), the class explores an example from that chapter that demonstrates the difficulty of obtaining accurate information on personal behaviors (p. 17): “Is a woman who has read in countless advertisements that nonbrushers are social offenders going to confess to a stranger that she does not brush her teeth regularly?” The randomized response method (Warner 1965) allows estimates of population parameters when “false” answers are included with known probabilities. This method not only makes the survey more ethical (at least in a Kantian sense), by giving respondents a measure (often quite strong) of privacy, but also makes the survey more accurate (for a reasonably healthy sample size) by reducing the respondent’s temptation to give inaccurate information. A method that offers a measure of privacy to respondents is of particular interest in an age where students encounter website cookies and linkable databases of financial and medical records. This is a natural opportunity to have students to look back at the ASA’s “Ethical Guidelines for Statistical Practice” (see Related Websites) to note the importance placed on privacy, such as in the section(II.D.) on responsibilities to research subjects. That section includes instructions to avoid excessive imposition, to obtain approvals from subjects appropriate for any secondary uses of data such as peer review, and to ensure privacy assurances take into account any legal limitations.

The method can be tailored to questions for which only “YES” is embarrassing (e.g., “Have you ever cheated in school?”), for which only “NO” is embarrassing (e.g., “Do you wash your hands after using the bathroom?”), or for which either answer could be controversial (e.g., “Do you favor making abortion illegal?”). Once the class is convinced that potentially embarrassing answers are indeed masked, students agree on a question and with great interest quickly collect and analyze the class grouped data. While Biblical analysis is beyond our scope, it is interesting to note that the idea of collecting information about people in an indirect manner out of respect has inspirations thousands of years old, such as when individuals are counted via the proxy of a coin donation (e.g., Exodus 30:13).

Randomized response is one of several places in the module where the detail of mathematics is easily adapted to the time available and/or mathematical background of the class, such as an accessible one-shot demonstration (Burger and Starbird 2000) or a more sustained, detailed analysis (Bolstad, Hunt and McWhirter 2001). Our module tended more towards the former, but yielded class reactions very similar to the authors of the latter paper (p. 149): “By the time we are ready to conduct the survey, they are fully satisfied that their privacy is being respected and are very interested in finding out what the class results will be. We find that almost everyone participates fully in the survey....Some students are quite emphatic that they would not mind giving their personal information, regardless of the randomization.” Before leaving this topic, it is important to make sure students do not think that the randomization involved within the randomized response method eliminates the need for randomization to select a valid sample of respondents in the first place!
A multiple-choice question on this topic from the module exam was: “Suppose a survey was conducted where 100 people each secretly flipped a coin. If they got heads, they were instructed to answer honestly the question: ‘Have you ever cheated on your current (or most recent, if now single) significant other?’ If they got tails, they were instructed to answer yes, regardless of the answer to the question. The data recording sheet showed 80 yeses and 20 noes. To the nearest tenths place, what is the best estimate of the proportion of people who have cheated on their significant other?” The answer choices were all multiples of .1 between 0 and 1.

3.2 Ethics of Displaying and Reporting Findings

Having addressed issues of how data is obtained, students must now be on guard for unethical ways data may be displayed, summarized, or reported. Students are exposed to typical examples (Huff 1993), occasionally updated with more recent examples (e.g., from CHANCE News; see Related Websites), but often discussion jumps to a less-routine level. For example, to let students explore the pitfalls of the types of average, the instructor gave the students an in-class group activity in which they were given a small number of “class sizes” (e.g., 10, 20, 30, 40, 105) and simply asked them to determine that university’s “average class size”. Depending on whether students use median or mean, and whether they average over class or over student, the answer to this question can range from 30 to 105! Hemenway (1982) shows why mean class size on a per-student basis (which seems more relevant from a student point of view) is never less than mean class size on a per-class basis.

During discussion, however, one student suggested discarding the “outlier” class size (105), an idea that triggered an unplanned, but insightful discussion of when one can justify discarding outliers. When to discard outliers was a focus of ASA Ethics Case Study #2 (see Related Websites). On an even larger scale is the “file drawer problem,” which concerns when it is ethical to ignore an entire study just because its findings fail to reach statistical significance. The Declaration of Helsinki actually states in principle #27: “Both authors and publishers have ethical obligations.... Negative as well as positive results should be published or otherwise publicly available.”

As another example, after discussing the straightforward distinction between norm-referenced (e.g., percentiles or a bell-curve) and criterion-referenced (e.g., explicit performance criteria not dependent on other students’ performance) assessment, the class then explores the ethical roots in how a teacher (or an institution) chooses and presents a particular balance between those two paradigms. A full-semester course could also explore the statistics and ethics associated with IQ tests, perhaps drawing from Krull and Pierce(1995) or some of the scholarly responses (Lipschütz-Yevick 1995) to the widely-publicized controversial work of Herrnstein and Murray (1994).

Finally, while it has been mentioned that simple “misuse” of statistics is usually viewed as an unsophisticated example of applied ethics, it should be noted that the difference between misuse and abuse is not always unambiguous. Students should discuss the caution of ASA (1999) guideline G3 to “[r]ecognize that differences of opinion and honest error do not constitute misconduct; they warrant discussion but not accusation. Questionable scientific practices may or may not constitute misconduct, depending on their nature and the definition of misconduct used.” Utts (2005, pp. 493-494) describes examples of the challenge of keeping politics from entering into decisions of how to collect or report federal data. With additional class time, it would also be interesting to explore the distinction Seltzer (2001) makes between the “traditional harm” of sloppy or distorted analysis and the “extraordinary harm” of using statistical and information systems to identify and attack individuals of vulnerable population subgroups (see Related Websites).

In broader society, there can be huge pitfalls in very basic phrases often taken for granted by teachers and authors of statistics that appear in the reporting of findings. As Schield (1995, p. 5) explains: “Students commonly use [‘predicts’, ‘accounts for’, ‘explains’] as causal; statisticians use most of these verbs
technically as indeterminate toward causality. This difference between common usage and technical usage creates a great opportunity for mistakes by students, for deception by those who are unethical opportunists and for silence by professionals who don’t see themselves as responsible for correcting the mistakes of others.” A Kantian would indeed see herself as responsible in this way, but an act-utilitarian might (out of necessity of having only 24 hours in a day) assess the cost-benefit ratio involved each time. A rule-utilitarian may suggest fighting the intentional misuses and any that are related to rules (but for other cases, fight as time permits). Students might be asked to discuss how they would respond to a new statistician’s scenario articulated by Vardeman and Morris (2003, p.23): “What looks to me like the thing that should be done would take two hours to explain and several more hours of my time to implement, while this client would be happy with something less appropriate that I could explain in five minutes.”

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 that Kantians would oppose. 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 data set in which one gender is favored at the aggregate level even though the other gender is favored at each sublevel (Lesser 2001a), a reversal known as Simpson’s Paradox. 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. Utilitarians, however, would 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 Gore-Bush 2000 Presidential campaigns and debates. The class considered a sequence of examples.

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 data set as if it represented two types of workers, regardless of company, such as educators and lawyers!). The support staff consist of 70 males each making $20,000 and 90 females each making $30,000. The executives consist of 30 males each making $90,000 and 10 females each making $100,000. Students can readily verify that females are paid better within each of the two categories, but less overall ($37,000 to $41,000). This type of data set gives students some context for discussing 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 [regardless of age, educational level, occupation, experience or working hours]...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 explored a data set of salaries in which all 5 males and 5 females are the same type of worker, and students verify that the women earn only 81 cents to the men’s dollar:

| WOMEN: Salary | $45200 | $28400 | $32000 | $34400 | $38000 |
|---------------|--------|--------|--------|--------|--------|
| Years of experience | 21 | 7 | 10 | 12 | 15 |
| MEN: Salary | $30800 | $42800 | $39200 | $52400 | $54800 |
| Years of experience | 9 | 19 | 16 | 27 | 29 |
When years of experience are then considered, students can verify that salaries can be explained perfectly by experience in this idealized example:

\[
PREDICTED \text{ SALARY} = 1200 \times (YEARS \text{ EXPERIENCE}) + 20000.
\]

While these particular numbers are simplified for classroom usage, it is true that women average fewer years of full-time paid work experience than men, and that this factor explains some of the earnings differential. As time and interest permits, students can explore actual data on these (and related) factors available from the Bureau of Labor Statistics and the Bureau of the Census (see Related Websites).

Another topical example of fairness in pay raises came from the current context of increasing schoolteacher accountability by tying raises to the learning (as measured by standardized test results) of their students. Students were asked to consider two teachers, Linda and Jane, whose students’ scores at year’s end were 95 and 70, respectively. Many students argued that it should depend on the improvement as well, at which point I uncovered the portion of my transparency that revealed that their students began the year at 85 and 50, respectively. On the one hand, 50 to 70 represents a larger gain, but on the other hand, on a bell curve distribution, it might not be a larger gain in standard deviations, and in any case, it is not even theoretically possible for a class beginning at 85 to improve 20 points due to the “ceiling effect.” This also led to discussion of teachers being tempted by such accountability measures to coach unduly or to “teach to the test”, the ethics of which have been discussed (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 are invoked, but it is done in a manner (Utts 1996) that need not require formal symbolism. Students are reminded of basic facts for probabilities and then given verbal, nontechnical explanations of the complement rule and of the multiplication rule for independent events. Paulos (1988, p. 20) offers one for the latter: “If...the outcome of one event has no influence on the outcome of the other, then the probability that they both occur is computed by multiplying the probabilities of the individual events.” There are specific court cases that have applied this law (Paulos 1988, p. 40), but the primary example of interest is the death penalty.

3.4.1 Death Penalty

The death penalty is a topic that has received a new wave of attention, thanks to the 2000 Presidential campaigns, prominent executions (e.g., Timothy McVeigh), and the work of the Innocence Project, and that offers an opportunity to revisit Simpson’s Paradox (Moore and McCabe 1999, p.207). Some defenders (Pambianco 2000) of the death penalty try to rebut “utilitarian arguments about fairness” by asserting that no innocent person has indeed been executed. We use the complement and multiplication rules to conduct a straightforward calculation that suggests the extreme implausibility of this claim.

To make this calculation more concrete and interactive, I ask the class for suggestions of the probability that an arbitrary death penalty case resulting in execution is indeed “the correct decision”. Students, whether from cynicism or awareness of the hundreds of cases of death row inmates exonerated by DNA evidence, typically volunteer numbers that range from .70 to .95. I then suggest we give the judicial system more benefit of the doubt and use .995. From the print or online versions of the 2000 Statistical Abstract of the United States (see Related Websites), we ascertained that during the period 1930-1999, there were 4458 executions by civil (not military) authorities that have been conducted in the United States. Raising .995 to the power 4458 suggests that the probability that all 4458 executions were correct decisions is a bit less than 1 in 5 billion. While the preceding analysis makes a very strong point, students discuss why there are some philosophical arguments for the death penalty that might be unaffected by this result. For example, a utilitarian could concede that innocents have been executed, but maintain that this
is outweighed by the number of innocent lives saved (e.g., by future crimes that may be prevented or deterred). One of the module exam questions assesses students’ ability to verify this type of calculation: “If the probability of correctness of a single decision is .96 and there are 9 people independently convicted for capital crimes, what is the probability (to the nearest tenths place) that at least one convicted person was wrongfully condemned?”

3.4.2 Profiling

Another topic from the 2000 Presidential debates and recent news (McAplin 2000, Kinsley 2001) is profiling. Because statistics is designed to gain information about groups more than about individuals, the issue of when it may be unethical to treat an individual as a member of group is of particular interest. To help students consider whether they should distinguish the case of when race is a component in a profile from the case of when it is the only component, and to generate big picture discussion (given heightened relevancy by airport security in the aftermath of September 11, 2001), the instructor has them read writers such as Kinsley (2001), who explains that the term racial profiling “has become virtually a synonym for racial discrimination. But if racial profiling means anything specific at all, it means rational discrimination: racial discrimination with a non-racist rationale.” He goes on to say that America has decided that such generalizations are morally wrong “even if they are statistically valid, and even if not acting on them imposes a real cost.” Kinsley attributes to Americans a type of Kantian position, which considers profiling immoral to the extent that it denies a person’s personhood or reduces a person to an object. Utilitarians, however, would view profiling as neutral or at least as an acceptable or necessary evil based on the “greatest good for the greatest number” criterion. One of the exam’s multiple choice questions was: “Profiling is something that would definitely be objected to by: a) utilitarians b) Kant c) all of the above d) none of the above”

In the case of motorists, 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 compare to the percentage of total miles driven by minorities or how they compare to the percentage of major accidents caused by minorities. Some textbooks (McClave and Sincich 2000) actually include (p. 136) exercises that report these facts that allow students to conclude that African-Americans were indeed “stopped more often for speeding than expected on this stretch of turnpike”: “14% of the drivers on this stretch of highway were African-Americans; 98% of the drivers were exceeding the speed limit by at least 5 mph (and, thus, subject to being stopped by the state police); of these violators, 15% of the drivers were African Americans; of the drivers stopped for speeding by the NJ state police, 35% were African American”.

Students also compare and contrast these types of profiling with the type of profiling routinely done by insurance companies, such as in using gender and age in setting auto insurance rates. Still another example to compare and contrast is the “successful student profile” that university admissions offices routinely construct based mainly on quantitative indicators (e.g., GPA and SAT/ACT). College students may be particularly interested in the case study (Utts 2005, pp. 113-116) in which a law based on a statistical profile of gender and drinking behavior for Oklahomans under 21 was challenged all the way to the U. S. Supreme Court in the 1970's and found discriminatory.

As a side note, students can explore how profiles could lead people to make conclusions that are logically impossible, such as (to use terms from the educational literature) when the representativeness heuristic leads to the conjunction fallacy (Utts 1996). Conceptually, we know P(A and B) cannot logically exceed P (A), but a student with the conjunction fallacy might think that the probability that ‘a randomly selected person is both an elementary teacher AND a woman’ could exceed the probability that ‘the randomly selected person is an elementary teacher’ if that student’s personal mental image of an elementary teacher is a woman.
3.5 Further Connections

Because a major distinction in utilitarian ethics is whether to apply the principle of utility to an individual situation (act-utilitarianism) or to a more general rule (rule-utilitarianism), there is a natural connection or analogy in statistics to have students encounter. Students can discuss how a statistician’s commitment to the Law of Large Numbers may relate to a rule-utilitarian’s commitment to “greatest good for the greatest number,” exploring the extent to which the latter is a big-picture, long run expected value that allows the possibility of an unfavorable outcome for that individual utilitarian in the short term. As Trowbridge (1989, p.11) states: “utilitarianism [is] very roughly stated as the greatest good for the greatest number over the longest period of time. Financial security systems rest on this same base....Whether the good that utilitarians attempt to maximize is called ‘happiness’, ‘pleasure’, or ‘utility’, and whether the maximization is individual or collective, are areas of controversy, but the general principle seems well accepted.” Because of the minimal mathematical background assumed, the Law of Large Numbers would be referenced only conceptually (Utts and Heckard 2004; p. 306), not in a mathematically technical way.

To give students immediacy, we offer a concrete scenario. Regardless of when one’s own birthday is, an act-utilitarian would agree that a populace would be better off overall if those with summer birthdays each had to pay $1000 and those with nonsummer birthdays each received $1000. In general, the “Is World A better than World B” type questions common to introductory philosophical courses on ethics or moral reasoning can usually be at least partially characterized by which distribution of wealth or happiness has the higher mean, smaller variance, etc.  [On a bigger picture level -- this could make a good exam or classroom discussion question -- students can decide if act-utilitarianism is off-limits to statisticians in light of the statement by Vardeman and Morris (2003, p.22): “Principled people consistently do principled work, regardless of whether it serves their short-term personal interests.”]

Students should be aware of limitations of these types of questions. For example, consider three worlds: world #1 is represented by a spinner with 100% chance of 50 units; world #2 is represented by a spinner with 1/3 chance of 30 units and a 2/3 chance of 60 units; world #3 is represented by a spinner with 3/5 chance of 40 units, 2/5 chance of 65 units.

![Diagram of a spinner with 50 units]
We see that all 3 worlds have the same expected value, so which of them might be regarded as “best” by utilitarians? Typically, most students declared world #1 most appealing to them (because all people would have the same amount) and viewed it most “utilitarian” as well (it would be the only world in which all members achieve that world’s greatest possible good). Other students said world #3 would better fit the utilitarian criterion because that world includes the highest possible greatest good (65). Still other students would prefer world #2, which has the highest median good (60) of all three worlds. So, the greatest minimum good (bringing to mind the maximin criterion from decision theory), greatest median good, and greatest maximum good would each choose a different world to fit the “greatest good for the greatest number” criterion. [For students not readily able to understand the spinners model, I also had a “finite population of fixed numbers” version in which the five people in world #1 each had 10 units of good; the five people in world #2 had 7, 7, 12, 12, and 12 units; the five people in world #3 had 8, 8, 8, 13 and 13 units.]

The primary point of the example has been made, but as a bonus, we can also explore pairwise comparisons of the worlds, finding that a resident in world #1 has a 2/3 chance of improving their lot by switching to world #2, a world #2 resident has a 3/5 chance of improving their lot by switching to world #3, and a world #3 resident has a 3/5 chance of improving their lot by switching to the original World #1. As one student put it, “no matter what world you choose, I can choose a world with the same average [mean] that would be better for most people,” which highlights an additional potential difficulty of implementing the utilitarian principle of “greatest happiness for greatest number.” While such potential of intransitivity is familiar (e.g., Efron’s dice) to statisticians, a similar situation can occur in the real world in the personal preference ordering of three or more alternatives (consider three individuals expressing these preferences among 3 options -- ABC, CAB, BCA, where a ballot of ABC means that voter considers option A her first choice, option B her second choice, and option C her third choice; we see that for the group, A is preferred to B in most of the ballots, B is preferred to C in most of the ballots, but A is not preferred to C in most of the ballots!). Another way to make the spinners example appear more realistic would be to assume the points refer to number of months of expected life (based on past medical data) under one of three available treatment options for a particular kind of cancer.

In real life, this question of which “greatest” (in the phrase “greatest good for the greatest number”) gets more weight also comes into play in the context of education in the face of statements such as NCTM (2000, p. 5): “All students should have the opportunity and the support necessary to learn significant mathematics with depth and understanding. There is no conflict between equity and excellence.”
class then discusses whether having 100% of students score at level 70 (on a 0-100 scale) is better than having 50% of students score at level 60, but 50% score at 90. A recent real-life example of this kind of question was posed by El Paso fifth-grade teacher Subramanian (2004): "Is it better to have a modicum of greatness surrounded by squalor or is it better to have no greatness at all where all the squalor has been eradicated?"

As time permits, decision rules (Lapin 1987) can be discussed other than the expected value, “utilitarian”-type approach such as the maximin (what’s best among the worst outcomes) or maximum likelihood (what’s best under the most likely outcome). Another pitfall is that even if there is a consensus to use expected value as the criterion, attitudes towards identical expected values can vary with whether the situation is framed as a potential gain or loss (Utts 1996, p. 227).

While Bentham’s quantitative form of utilitarianism yielded to Mill’s more qualitative form (see Appendix B), it is interesting to examine how structuring stochastic situations can provide mechanisms to indirectly quantify human values that many believe cannot (or, in some cases, should not) be quantified. For example, Gelman (1998, p. 172) maintains that people “really do put money and lives on a common scale, whether they like it or not. There is a vast literature on the practical, political and moral issues involved in equating dollars and lives; see, for example Rhoads 1980 and Dorman 1996.” When Gelman asks his university students how much money would they accept in exchange for being killed, they generally answer that they would not be killed for any amount of money. Gelman later tests consistency by asking students whether they would prefer (a) their current situation or (b) a small probability p of dying and a probability 1-p of gaining $1000, and notes that “to see that $1000 is worth a nonnegligible fraction of a life, consider that people will not necessarily spend that much for air bags for their cars.”

The tension of applying statistical approaches to human problems is articulated by Paulos (1991, p. 67): “Balancing cost vs. value in medical care or price vs. impact in environmental protection is always an unpleasant task. There are times, however, when not being quantitative is a kind of false piety which can only make obscure and thus more difficult the choices we must make.” Paulos continues by invoking a more Kantian direction: “There are also other times, I hasten to append, when the economic arithmetic that is appropriate is more Cantorial..., when every life is of infinite worth and thus not less valuable than the sum of any number of other infinitely valuable lives.” An aim of the module is to apply statistics to real-world scenarios to show (1) how useful and necessary quantitative approaches can be in exploring them and (2) the challenges of finding precise implementations or interpretations of common ethical principles.

4. Challenges and Lessons

4.1 Emphasis on Engagement

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. (This brings to mind the spirit of aphorist Mason Cooley’s remark that “Reading about ethics is about as likely to improve one’s behavior as reading about sports is to make one into an athlete.”) 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.” (The number of students completing evaluations represents roughly 80% of the students who completed the author’s module; some of the 150 enrolled in the course opted to take the course for 2 credit-hours rather than 3, and therefore were obligated to complete one less module than the 3-credit hour students. Though collected after each module, the formal student evaluation data was not made available until the entire semester was over. Nevertheless, each professor delivering a module 4
consecutive times certainly had the fairly rare opportunity to refine the module during the semester based on informal class response each iteration.)

Not only do students profit by having discussion as well as lecture, there is also benefit to having assessment items that are short-answer or essays about scenarios, rather than just lower-level multiple-choice items. As Ozar (2001, p. 22) notes, “Ethics teaching is most effective and the learning outcomes seem to be most effectively secured in the learners when they are required to write case-based essays about the issues they have been studying.” Case studies and ideas for case studies can be found in many places (see Related Websites). Even faculty battling severe time constraints can at least make regular assignments of “minute-papers” or log entries for students. Those faculty who may avoid making writing assignments out of a lack of confidence to evaluate them properly (as Ozar 2001 maintains) will appreciate knowing that examples of grading rubrics and criteria for such assignments are offered by Newton (2001), who also notes that field-specific applied ethics modules can utilize a fuller set of criteria than a general ethics class.

4.2 Addressing Background and Textbook Limitations

As mentioned earlier, a 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. The instructor felt comfortable in this area (as did Webber 1997) after doing some reading on his own and consulting a local philosophy professor (who later became the second author of this article) as a resource. However, there were students (especially those who had to miss one meeting, or 20% of the non-exam part of the module!) who were not used to having 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 module as he would have if students had more self-contained resources to fall back on to supplement class meetings.

Because few of the students in the module had taken an introductory course in statistics (although the author hoped more would be inspired to do so, of course), there was a need to find a way to present the information with as little technical jargon or formal symbolism as possible. Books such as Haack (1979) and Utts (1996) gave the instructor ideas for meeting this challenge.

4.3 Moving Beyond Superficial Absolutism and Relativism

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, etc.), thus illustrating the importance of making one’s definitions, values, and assumptions explicit. A concrete set of striking demonstrations of this importance was recently offered by Melton (2004). At times we found, as did Oberhoff and Barnes (2000, p. 50), that our examples “generated considerable student discussion,” including whether there always existed a ‘best solution’ to an ethical problem (compare with ‘best interpretation’ to a data set).

By the time students are in college, though, they seem more likely to overestimate rather than underestimate the fallibility or relativism of statistical analysis, which might also be understood in terms of the literature on cognitive moral development (e.g., Perry 1970, who describes this progression: dualism, multiplicity, relativism, committed relativism). It seemed critical, therefore, to emphasize that the apparent ambiguity of some scenarios does not lock us into a relativist approach because our ability to reason with statistics can identify theoretical preferences and comparisons and definitively rule out inferior or unreasonable solutions. For example, while there is certain sensitive personal information that we cannot
expect to obtain directly and accurately from individuals, we can do much better than each of us making a subjective guess now that we have learned about the tool of randomized response.

The instructor found it helpful to consult Momeyer (1995), who explores and reflects on many different forms of, and motivations for, relativism by students and teaching strategies of how to address them. Momeyer’s strategies include treating students respectfully as open-minded skeptics sincerely seeking reasonable solutions, being willing to share one’s own convictions, and allowing students to debate, work with, and learn from their peers. Being aware of how to address student relativism helped the author heed Bok’s warning (1976, p. 30) 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.”

4.4 Student Evaluations of the Module

The student feedback on the course was very positive. On the narrative part of AASU’s formal end-of-course evaluations, many students expressed satisfaction at just how much the module managed to offer connections. This even included students who had poor attitudes towards the content area prior to the course, as indicated by written comments such as: “When I saw it on the syllabus I thought it was going to be boring and hard. But it actually turned out very nice.” and “I hate math and must say this opened my eyes to a whole different view of how math is part of our everyday lives.” Open-ended written evaluation feedback consistently described the class as “very enjoyable”, related to everyday life, and revealing pitfalls previously unknown to the student. On the quantitative portion of the evaluations (using a 1-2-3-4 scale, where 4 is the best rating on each of the 11 quantitative questions), the mean of the student evaluations of the author’s module was 3.47 (N=81 responses) for the 2000-2001 school year and 3.64 (N=83) for the 2001-2002 school year. The latter year’s rating included only 2% “2” responses and no “1” responses at all out of the 913 total responses for all students on all items.

4.5 Concluding Thoughts

Universities should find statistics to be one of the most rich and universal vehicles for implementing an interdisciplinary ethics requirement. As we have discussed, statistics has the capacity to include any appropriate quantification and has the commitment to examine critically all types of assumptions.

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. Shulman (2002), however, indicates that a separate course may not be desirable even when possible, and makes a strong argument for integrating ethics directly into already existing required mathematics and science classes. For example, University of Texas at El Paso Professor Joan Staniswalis incorporates the following assignment into her introductory statistics course: "Go to the website www.nlm.nih.gov/pubs.cbm/hum_exp.html, click on Table of Contents, then click on Historical Perspectives on Research Involving Human Participants. Find one article, read it, and write a one paragraph summary. Make sure that the complete reference is listed on your summary." The importance of the topic will likely be impressed upon the students by this bibliography's size alone!

The author’s awareness of the dangers of teaching content without sufficient emphasis on accompanying ethical issues or responsibility dates back to his very first semester of teaching statistics (a business statistics course in 1988) in which an entire lesson of pitfalls of descriptive graphical summaries was followed by a student’s (unfortunately, serious) question: “So is this [as the student is pointing to an example of a misleading graph] how we’re supposed to do it when we’re working in the real world?” We hope we have outlined a rich sampling of meaningful resources and connections between statistics and
Appendix A: Kantian Ethics

Students are first introduced to Kant’s ethical theory and the universalizability test contained within his categorical imperative. Kant sets out to isolate specifically moral worth as opposed to other kinds of worth. There are various kinds of reasons that we might respect or value people and their actions, including their skills, talents, achievements, good fortune, fame, and natural endowments, but none has anything to do with moral worth. We might admire Hitler as a military strategist, but we regard him as an evil man. The respect we have for his particular skill is certainly not moral respect. So moral worth does not consist in having a particular skill or talent. Kant states that the only basis of moral worth is a good will. Good will does not mean having a sunny disposition or being nice. It means: a will that is capable of acting from duty and our duty is ultimately determined by a rational and universal standard, so we can also say that a good will is one that obeys the dictates of reason.

The central question then becomes: What is our duty? Kant proposes a superrule, called the categorical imperative to answer this question. A categorical imperative simply says: “do this!” or “don't do this!” without any conditions. “Keep your promises” and “don't steal” are categorical. They contrast with hypothetical imperatives, which are "if...then" imperatives, and advise us about the best means to achieve a given end, but do not tell us what ends to seek. An example: if you want to be a lawyer, go to law school. But also: if you want to be a burglar, let an expert thief teach you. The point is: hypothetical imperatives do not tell us what ends we should or should not seek, whereas for Kant any rational ethic must tell us whether the ends are right or wrong. For this reason Kant insisted that most past ethics was mistaken because it rested on a hypothetical imperative of the following kind: If you want to be happy, don't steal! And he objected to this for two reasons: (1) attaining happiness is not the proper motive for doing our duty, and (2) as a matter of fact, it is simply false to say that not stealing or doing right will make a person happy. After all, a person's happiness depends upon all sorts of subjective factors that differ from person to person, and Kant wants to insist that what is right or wrong is an objective matter. The categorical imperative is intended to give us a universal objective moral standard!

The main formulation of the categorical imperative is: Always act so that the maxim of your action can be a universal law. This command is designed to show us that when we try to universalize certain actions they become self-defeating. Suppose we are tempted to cheat at cards. The maxim would be something like: cheat when you can get away with it. Not only will this not pass the golden rule test, but more importantly, if everyone acted on this maxim we would no longer know what cheating meant, since games are defined by rules and cheating cannot occur unless there are rules. If everyone is cheating, there are no rules and hence no game. And where there is no game it is not possible to cheat. The point of my cheating, in short, depends on everyone else following the rules. Similar points can be made about breaking promises, lying, and other human activities.

Weaknesses of Kant’s view, like the inability to take account of consequences in the assessment of duty, are also observed. Whatever the defects of his view, Kant's claim that there is a distinctively moral worth that we recognize and that there is a distinctively moral standpoint is hard to get away from. In particular, his claim that when we make moral judgments (claims about right and wrong, fairness and unfairness, etc.) that we are implicitly claiming to occupy a disinterested, umpire-like standpoint is very hard to get around. In fact, he thinks this power to act from duty even when it means ignoring our own interests is what makes us distinctively human.

Appendix B: Utilitarian Ethics
Utilitarianism is the name for a family of ethical theories that were first developed in the nineteenth century. Its founders were Jeremy Bentham and John Stuart Mill. All versions of utilitarianism are teleological, they define what is morally right as that which produces the most good (or the least bad) possible under the circumstances. They differ from Kant, then, in saying that consequences, not just intentions, determine moral right and wrong. Utilitarians have their own superrule called the principle of utility which proclaims: the greatest good of the greatest number!

In the early nineteenth century, Jeremy Bentham wanted to reform English institutions. Bentham (1988) begins with the thesis called psychological egoistic hedonism: that people are always motivated by a desire to secure pleasure and avoid pain. So he defined the good as pleasure and then proposed the principle of utility as the best way to help people attain this goal. He thought that by using this principle and adding up the likely pleasures and pains that would result from any action or proposed legislation, the human race would make much more rational decisions. He even proposed a pleasure-calculus according to which we should weigh pleasures and pains according to the following factors:

1. How intense is the pleasure?
2. How long will it last?
3. How certain is it to occur?
4. How pure is it or will it lead to later pain?
5. Does it lead to other pleasures?
6. Is it near or far off in time?
7. How many people are affected by it and how?
   (each person counts equally)

The development of the pleasure-calculus is historically significant because it clearly establishes a model of how mathematics should be used in utilitarian ethical theory. With the quantification of pleasure and calculation of the numbers affected, the moral correct course of action would suddenly appear mathematically self-evident. Utilitarianism, in all of the forms it will take and in contrast to Kantianism and virtue ethics, affirms its direct relation to and need for statistics in any reflective attempt to discover what is moral. How could the “greatest number” ever be calculated without statistics playing a central role in morality? The qualitative assessment of “greatest good” becomes inextricably bound, in both theory and application, to the quantitative calculation of “greatest number” in utilitarianism. It is worth noting that as utilitarianism evolves, the word “good” in the principle of utility becomes increasingly harder to define while the phrase “greatest number” becomes increasingly more accurate and easier to compute. As utilitarianism developed from Bentham’s quantitative hedonism (where good = quantity of pleasure) to Mill’s qualitative hedonism (where good = the qualitatively higher intellectual pleasures outranking the lower physical pleasures) and eventually to Ideal utilitarianism (where the word “good” can no longer be clearly or specifically defined), the increasing clarity, specificity, and accuracy of modern statistical calculations of the “greatest number” offset the theoretical disadvantages of not defining the word “good” in principle of utility.

The debate between act- and rule-utilitarians does not affect the aforementioned point since they are essentially engaged in a statistical debate regarding what counts as the “greatest number” in a given case and what benefits the “greatest number” more — a specific act or adherence to a rule. For example, in the often cited lifeboat example where there are ten spaces in the boat for eleven people, act-utilitarianism would morally authorize letting someone die as producing the greatest good for the greatest number. And it would, in contrast to Kant, allow us to break promises, lie, kill, etc., when a balance of good over bad is produced by doing so.

However, this advantage is also a drawback. For in some cases the principle of utility would seem to declare actions to be morally right that our deepest moral instincts tell us are morally wrong. Consider the following case: In a small Southern, racist town a white woman has been killed. The police have no clue
as to the murderer. But the Klan is about to burn a bunch of black houses and begin a reign a terror in which a lot of people will be hurt or killed. So the police randomly pick out a black man whom they know to be innocent and fabricate evidence against him. He is tried and sentenced and hanged. This quiets the Klan and eliminates the threat of further killings and violence. Hanging an innocent man (or in other cases, letting a guilty man go free) prevents even more death and destruction and thus serves the greatest good of the greatest number. Therefore, it is, according to the principle of utility, the right thing to do! But this conflicts with our ordinary sense of right and wrong. Can a theory that is so much out of sync with our common sense be correct? Most people do not think so. In fact, many utilitarians are themselves unwilling to accept this, so they have tried to fix the theory and eliminate this kind of embarrassing result. They do so by moving to a position called rule-utilitarianism.

Modern defenders of utilitarianism try to improve the theory by making a rather ingenious move. They claim that we should not apply the principle of utility directly to individual acts, i.e., on a case by case basis, but only to the rules, laws, and practices that prevail in a society. They point out that societies have rules, laws, and practices that define and regulate individual actions. These evolve or are deliberately changed from time to time, but as long as they are in force, the individual should follow them. For the rule-utilitarian, the proper question to ask is: does this rule, practice, or law serve the greatest good of the greatest number on the whole? That is, if it were absent, would society be worse off? If it would be, then the rule should not be compromised in an individual case.

In other words, the phrase “greatest number” for the rule-utilitarian must always refer to society as a whole and society over time rather than the majority of individuals directly affected by the act. Rule-utilitarianism is in a much stronger position to meet the kind of moral objections raised by hanging the innocent man. Its view is that the rule against hanging innocent people serves human society sufficiently well on the whole so that in cases were more local good results from giving it up, we are still not entitled to do so.

The act-utilitarian, on the other hand, is clearly more radical and will often make decisions that conflict with our ordinary moral intuitions. He can only weakly defend rules rough guidelines (“rules of thumb”) that may be violated when they do not serve the greatest good of the greatest number (which here equals the largest number of people involved in the case at that point in time). And the act-utilitarian has a more difficult time assessing consequences, since he deals with individual actions case by case, whose effects are tougher to reasonably predict than those of social practices with which the human race has long experience and have stood the test of generations.

**Appendix C: Virtue Ethics**

Discussion of virtue ethics, as derived from Aristotle (1980), serves as a means of illustrating just how far modern ethical theories have integrated rule-bound quantitative methods into their understanding of the “good”. Essentially, the virtue ethics theory differs from the aforementioned theories in that it does not depend on consequences, feelings, pleasure, or most importantly, rules so much as on human beings developing a moral or virtuous character by doing what a good or “virtuous” person would do. Instead of focusing on rules or rule-following as Kant does with his superrule, the categorical imperative, or as the utilitarians do with their principle of utility, Aristotle sees ethics and moral activity as being dependent on a person’s character which takes years to develop and is the only way to true human fulfillment (i.e., “Eudaimonia”, sometimes translated as “happiness”). If a person is not virtuous, no amount of rule following, or statistical assessment, will ever allow that person to consistently be good and act well, but more importantly such a person (i.e. one who depends on rules in order to be good) can never be “happy” according to Aristotle because they have not developed that specific function of a human being that makes them human--namely the ability to reason well or as Aristotle puts it to “reason in accordance with virtue”.
Aristotle (1980, p. 13) defines virtue as “that state of a thing which constitutes its peculiar excellence and enables it to perform its function well ... in man [it is] the activity of reason and of rationally ordered habits.” The emphasis is on the good or virtuous character of human beings themselves, rather than their acts, the consequences of their acts, methods of calculation, feelings or rules. In other words, it is the development of the good person that is important in this moral theory, not abstract rules, or consequences of acts or rules except as they derive from a good or virtuous person or cause that person to become good or virtuous.

The remainder of the Aristotle’s book is devoted to explaining what it means to “reason well”. It has numerous details and distinctions such as the division of reason into two main parts: intellectual reasoning vs. moral reasoning. Each part has its own virtues, that is, its own excellences. The most suggestive detail relevant to reflections on the connections between ethics and statistics is Aristotle’s understanding of moral reasoning that always aims at a mean between extremes (i.e., The Golden Mean), thus focusing on moderation as an essential element of the virtuous person. The use of the word “mean” in an quasi-mathematical sense to define moral virtue is worthy of considerable speculation. At the very least, Aristotle appears to intimate that those completely without mathematical sensibilities as an element of their reasoning will be unable to be virtuous.

Acknowledgments

This work was presented in part at the 2001 Joint Statistical Meetings (Lesser 2001b) and was supported in part by the first author’s Arthur M. Gignilliat, Jr. Professorship at Armstrong Atlantic State University, for which this author expresses gratitude. Also, the authors thank Tom Short and the anonymous referees and associate editors for particularly helpful responses that resulted in significant improvements to this paper.

References

Advisory Committee to the Surgeon General of the Public Health Service (1964), Smoking and Health. Public Health Service publication no. 1103, Washington, DC: U.S. Government Printing Office. Also available on the world wide web at: [www.cdc.gov/tobacco/sgr/sgr_1964/sgr64.htm](http://www.cdc.gov/tobacco/sgr/sgr_1964/sgr64.htm)

Annenberg/CPB (1989), “The Question of Causation; Experimental Design”. Videos 11 and 12 of the 26-video series Against All Odds: Inside Statistics.

Arbuthnot, J. B. and Faust, D. (1981), Teaching Moral Reasoning: Theory and Practice, New York: Harper and Row.

Aristotle (1980), The Nicomachean Ethics [Translated from the Ancient Greek with an introduction by D. Ross. Revised by J. L. Ackrill and J. O. Urmson], Oxford, England: Oxford University Press.

Bentham, J. (1988), Introduction to the Principles of Morals and Legislation (originally printed in 1781 and published in 1789), Amherst, New York: Prometheus Books.

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

Bolstad, W. M., Hunt, L. A., and McWhirter, J. L. (May 2001), "Sex, Drugs, and Rock & Roll Survey in a First-Year Service Course in Statistics," The American Statistician, 55(2), 145-149.
Bradford-Hill, A. (1965), “The Environment and Disease: Association or Causation?” President’s Address, Proceedings of the Royal Society of Medicine, 9, 295-300.

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

Cook, R. D. (1980), “Smoking and Lung Cancer,” in R.A. Fisher: An Appreciation, Lecture notes in statistics, eds. Fienberg, S. E. and Hinkley, D. V., Vol. 1, pp. 182-191, New York: Springer-Verlag.

Doll, R. and Hill, A. B. (1950), “Smoking and Carcinoma of the Lung,” British Medical Journal, 2, 739-748.

Doll, R. and Hill, A. B. (1952), “A Study of the Aetiology of Carcinoma of the Lung,” British Medical Journal, 2, 1272-1286.

Doll, R. and Hill, A. B. (1956), “Lung Cancer and Other Causes of Death in Relation to Smoking,” British Medical Journal, 2, 1071-1081.

Dorman, P. (1996), Markets and Mortality: Economics, Dangerous Work, and Value of Human Life, Cambridge: Cambridge University Press.

Freedman, D. (1999), “From Association to Causation: Some Remarks on the History of Statistics,” Statistical Science, 14(3), 243-258.

Frosch, R. (December 2001), “Presidential Invited Address at JSM,” Amstat News, No. 294, 7-16.

Gardenier, J. S. (February 2002), “Making Statistical Ethics Work for You, of Course,” Amstat News, No. 296, pp. 21-22.

Gelman, A. (1998), “Some Class-Participation Demonstrations for Decision Theory and Bayesian Statistics,” The American Statistician, 52(2), 167-174.

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

Griffith, W. B. (April 1995), “Moral Codes and Professional Ethics.” Amstat News, No. 219, 13-14.

Haack, D. G. (1979), Statistical Literacy: A Guide to Interpretation, North Scituate, MA: Duxbury.

Heller, J. (July 26, 1972), “Syphilis Victims in U.S. Study Went Untreated for 40 Years,” New York Times, p.1,8.

Hemenway, D. (1982), “Why Your Classes are Larger than ‘Average,’” Mathematics Magazine, 55(3), 162-164.

Herrnstein, R. J. and Murray, C. A. (1994), The Bell Curve: Intelligence and Class Structure in American Life, New York: Free Press.

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

Jones, J. H. (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 News*, p. 9A.

Krull, C. D. and Pierce, W. D. (1995), “IQ Testing in America: A Victim of its own success,” *The Alberta Journal of Educational Research*, 41 (3), 349-354.

Lapin, L. L. (1987), *Statistics for Modern Business Decisions (4th ed.*)*, Austin: Harcourt Brace Jovanovich.

Lesser, L. M. (2001a), “Representations of Reversal: An Exploration of Simpson's Paradox,” in *The Roles of Representation in School Mathematics*, eds. Albert A. Cuoco and Frances R. Curcio, pp. 129-145, Reston, VA: National Council of Teachers of Mathematics.

Lesser, L.M. (2001b), "Ethical Statistics and Statistical Ethics: The Experience of Creating an Interdisciplinary Module," *2001 Proceedings of the American Statistical Association Section on Statistical Education [CD-ROM]*, Alexandria, VA: American Statistical Association.

Lipschütz-Yevick, M. (1995), “The Questionable Probability Theory Behind the Strange Story of The Bell Curve’s Bell Curve,” *Humanistic Mathematics Network Journal*, 12, 22-23.

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

McClave, J. T. and Sincich, T. (2000), *Statistics (8th ed.*)*, Upper Saddle River, NJ: Prentice Hall.

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

Melton, K. I. (2004), “Statistical thinking activities: Some simple exercises with powerful lesson,” *Journal of Statistics Education [Online]*, 12(2). [ww2.amstat.org/publications/jse/v12n2/melton.html](http://ww2.amstat.org/publications/jse/v12n2/melton.html)

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

Miller, F. G. (2002), “Ethical Significance of Ethics-Related Empirical Research,” *Journal of the National Cancer Institute*, 94(24), 1821-1822.

Moe, K. (1984), “Should the Nazi Research Data Be Cited?” *Hastings Center Report*, 14 (6), 5-7.

Momeyer, R. W. (1995), “Teaching Ethics to Student Relativists,” *Teaching Philosophy*, 18(4), 301-311.

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

Moore, D. S. (2001), *Statistics: Concepts and Controversies*, New York: W. H. Freeman.

National Council of Teachers of Mathematics (2000), *Principles and Standards for School Mathematics*, Reston, VA: NCTM.

Newton, L. (2001), “Outcomes Assessment of an Ethics Program,” *Teaching Ethics*, 2(1), 29-67.
Oberhoff, K. and Barnes, R. (September 2000), “Senior Seminar: A Capstone Course in the Computer and Mathematical Sciences,” *Humanistic Mathematics Network Journal*, 23, 49-54.

Ozar, D. T. (2001), “Ethics Across the Curriculum Programs,” *Teaching Ethics*, 2(1), 1-27.

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

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

Paulos, J. A. (1991), *Beyond Numeracy: Ruminations of a Numbers Man*, New York: Vintage Books.

Perry, W. (1970), *Forms of Intellectual and Ethical Development in the College Years*, New York: Rinehart and Winston.

Rhoads, S.E. (1980), *Valuing Life: Public Policy Dilemmas*, Boulder, CO: Westview Press.

Rose, M. M. (1996), “Integrating Ethics and Diversity into Economics and Business Statistics,” *Journal of Education for Business*, 72(1), 13-17.

Schield, M. (1995), “Correlation, Determination, and Causality in Introductory Statistics,” *Proceedings of the ASA Section of Statistical Education*, 189-194, Alexandria, VA: American Statistical Association.

Schield, M. (1999), "Simpson's Paradox and Cornfield's Conditions," *Proceedings of the ASA Section of Statistical Education*, 106-111, Alexandria, VA: American Statistical Association.

Seltzer, W. (2001), “U.S. Federal Statistics and Statistical Ethics: The Role of the American Statistical Association’s Ethical Guidelines for Statistical Practice,” Paper presented at Methodology section seminar, Washington Statistical Society. Available on the world wide web at: www.uwm.edu/%7Emargo/govstat/wss.pdf.

Shulman, B. (2002), “Is There Enough Poison Gas to Kill the City?: The Teaching of Ethics in Mathematics Classes,” *College Mathematics Journal*, 33 (2), 118-125.

Sia, S. (2001), "Teaching Ethics in a Core Curriculum: Some Observations," *Teaching Ethics*, 2(1), 69-75.

Subramanian, R. (October 21, 2004), "No Child Left Behind Act Benefits Educators," *El Paso Times*, 124 (295), p. 5B.

Thiroux, J. P. (2001), *Ethics: Theory and Practice*, Upper Saddle River, NJ: Prentice Hall.

Trowbridge, C. L. (1989), *Fundamental Concepts of Actuarial Science*, Schaumburg, IL: Actuarial Education and Research Fund.

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

Utts, J. M. and Heckard, R. F. (2004), *Mind on Statistics (2nd ed.)*, Belmont, CA: Brooks/Cole.

Utts, J. M. (2005), *Seeing Through Statistics (3rd ed.)*, Pacific Groves, CA: Brooks/Cole.
Vardeman, S. B. and Morris, M. D. (2003), “Statistics and Ethics: Some Advice for Young Statisticians,” *The American Statistician*, 57(1), 21-26.

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*, 60 (309), 63-69.

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

Wynder, E. L., Graham, E. A., and Croninger, A. B. (1953), ”Experimental Production of Carcinoma with Cigarette Tar,” Part I, *Cancer Research*, 13, 855.

Wynder, E. L., Graham, E. A., and Croninger, A. B. (1955), ”Experimental Production of Carcinoma with Cigarette Tar,” Part II, *Cancer Research*, 15, 445.

**Related Websites**

American Statistical Association (1999), “Ethical Guidelines for Statistical Practice.”
[ww2.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics](http://ww2.amstat.org/profession/index.cfm?fuseaction=ethicalstatistics)

Roger Boisjoly
[www.onlineethics.org/codes/index.html](http://www.onlineethics.org/codes/index.html)

United States Holocaust Memorial Museum
[www.ushmm.org/](http://www.ushmm.org/)
Nazi experiments
[www.ushmm.org/research/doctors/twoa.htm](http://www.ushmm.org/research/doctors/twoa.htm)
Nuremberg Code
[www.ushmm.org/research/doctors/Nuremberg_Code.htm](http://www.ushmm.org/research/doctors/Nuremberg_Code.htm)

International Statistical Institute (1985),"Declaration of Professional Ethics for Statisticians." [www.cbs.nl/isi/ethics.htm](http://www.cbs.nl/isi/ethics.htm)
(apparently also in (1986) *International Statistical Review*, 227-247.)

United Nations statistical division (1994), “Fundamental Principles of Official Statistics.”
[unstats.un.org/unsd/statcom/doc94/e1994.htm](http://unstats.un.org/unsd/statcom/doc94/e1994.htm)

The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research
[www.fda.gov/ohrfrb/humansubjects/guidance/belmont.html](http://www.fda.gov/ohrfrb/humansubjects/guidance/belmont.html)
[www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm](http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm)

Declaration of Helsinki
[http://www.wma.net/e/policy/b3.htm](http://www.wma.net/e/policy/b3.htm)

CHANCE News database
[www.dartmouth.edu/~chance/chance_news/news.html](http://www.dartmouth.edu/~chance/chance_news/news.html)

ASA Ethics Cases
[www.TCNJEUDU/~ethcstat/cases.html](http://www.TCNJEUDU/~ethcstat/cases.html)
Papers (by Seltzer, Anderson, etc.) on confidentiality ethics:
www.uwm.edu/~margo/govstat/integrity.htm

Bureau of Labor Statistics
stats.bls.gov

Bureau of the Census
www.census.gov

Statistical Abstract of the United States
www2.census.gov/prod2/statcomp/index.htm

Lawrence M. Lesser
Department of Mathematical Sciences
University of Texas at El Paso
500 W. University Ave.
El Paso, TX 79968-0514
U. S. A.
Lesser@utep.edu

Erik Nordenhaug
Department of Languages, Literature and Philosophy
Armstrong Atlantic State University
11935 Abercorn Street
Savannah, GA 31419
U. S. A.
nordener@mail.armstrong.edu

Volume 12 (2004) | Archive | Index | Data Archive | Information Service | Editorial Board | Guidelines for Authors | Guidelines for Data Contributors | Home Page | Contact JSE | ASA Publications