Training University Students in Ethical Use of Statistics as Part of the LANGURE Project

Larry Nelson
Statistics, North Carolina State University, Campus Box 8203, Raleigh, NC, 27695

Training University Students in Ethical Use of Statistics as Part of the LANGURE Project

Abstract

The need for an Ethics course gave rise to a project called LANGURE involving a consortium of eight US universities. The resulting one-hour course will be taught at several Land Grant and related institutions. In some institutions, it will be required of all graduate students. One important course module will be Ethical Use of Statistics. Statisticians and users of statistics from the eight universities are jointly involved in developing the content of the course. This content will focus on types of ethical problems in statistical usage such as misuse, sloppiness and outright fraud. Case study examples of each of these problem types will be presented in the course. In addition to planning and conducting experiments or surveys, emphasis will also be given to the analysis and reporting phases. Module access will be available on the NCSU Graduate School website.

KEY WORDS: Ethical Use of Statistics, LANGURE Project, Research Ethics Course

Larry A. Nelson, Professor Emeritus of Statistics
North Carolina State University

1. LANGURE Ethics in Research Course

1.1 Introduction

There has been an increased interest globally in Ethics as applied to many phases of life.
One doesn't need to take time to give examples of recent breaches in Ethics at the state or national levels. A number of universities are now emphasizing training in Ethics either in a formal way or in more informal ways such as seminars and special lectures. Teaching Ethics is necessary in the university academic setting as cheating, plagiarism and related unethical behaviors are frequently encountered in classroom activities. The focus of the LANGURE Project is more on the application of Ethics to Research and the focus of this paper is more specifically on the Ethical Use of Statistics in Research. The LANGURE acronym refers to "Land Grant University Research Ethics". The LANGURE be a required course for all Ph.D. students.

1.2 Structure of Consortium and Resulting Course Development

The course is rather free-formatted, i. e. there will be a core course in a classroom lecture of two hours a week for seven weeks (or one hour a week for 14 weeks) but there has also been a free on-line generic version of the course called the OPEN SEMINAR IN RESEARCH ETHICS developed. An instructor could modify the content of either the classroom or on-line versions to suit his or her local needs and the two versions or portions thereof could be taught simultaneously (students would study the on-line version before going to the more formal lecture). Much of the course is devoted to lectures on the ethical theory and only about 1/7 of the total time will be spent on one of the various modules. For the two-hour a week course, the seventh week will be spent on a module. These will be taught in the student's home department.

The Statistics module for the OPEN SEMINAR IN RESEARCH ETHICS is located at http://www.Chass.ncsu.edu/langure/modules/statistics.html

A National Science Foundation Grant of $250,000 for a period of three years was awarded to the consortium in 2005 to develop the one-hour graduate course in Research Ethics, which could be taught at those universities (and/or others). NSF staff was interested in the Land Grant Institutions because many of them do not have medical schools so students do not have access to the Ethics training which government agencies supported for medical students' ethics training. The universities are as follows: North Carolina State University, University of Wisconsin, Iowa State University, University of Hawaii, North Carolina Central University,

North Carolina A. and T. State University, Purdue University and Fayetteville State University. An institutional leader was named at each university to coordinate the overall program at that university. As part of the new Ethics in Research Course, there were a number of modules developed according to the subject matter areas. The Lead Institution, North Carolina State University in Raleigh, named a Senior Fellow and a Junior Fellow (usually a graduate student) to lead the development of each module. I was the Senior Fellow for the Ethical Use of Statistics Module and Dr. Michael Crotty was the Junior Fellow when he was a graduate student. Also each cooperating institution could name a Senior Fellow for each module. In addition to the Statistics focus discipline, some of the disciplinary modules developed were Agricultural Biotechnology and Sustainability, Biomedical Engineering, Computer Science, Ethical Business Communication, Field Ecology, and Use of Animals for Food and Fiber. A number of other modules were developed. For the on-line version, the administrators of the program decided that the various modules would more or less be stand-alone and the students could negotiate them separately on the web. There was no effort to standardize their major topics or format.

A national (and to a lesser extent international) meeting was held in Raleigh, North Carolina in April of 2007 to discuss the overall effort and to develop plans for the remainder of the project. Experts exchanged ideas and learned how to teach the free online course, OPEN SEMINAR IN RESEARCH ETHICS. This on-line course will also be available to students in other universities, which are not part of the LANGURE consortium (including universities outside the U.S.).

1.3 Description of LANGURE Approach

The LANGURE approach is different from the traditional Responsible Conduct of Research (RCR) approach, which emphasizes compliance with a set of rules and regulations. What might be considered ethical practice may vary from one application area to another. Critical questioning is the LANGURE approach to learning Ethics. The paradigm used is called Research Ethics Education (REE). Critical inquiry is nurtured and students are welcomed into the community of scholars and entrusted with a sense of community responsibility inherent in being a part of that community. They are taught to apply reasoning to each situation in which ethics is involved and consider various courses of action. General areas where problems might occur are pointed out. Case studies are used to generate thinking and discussion, and then students' reactions to the case studies follow.

2. Ethical Use of Statistics Module

2.1 Professional Audience

Early on, it was necessary to decide what was the target audience for the Statistics module. Statistics majors would be part of the graduate student population taking the module, but the number of students in other majors would be much greater. Initially, time was not available to develop two separate modules for Statistics, including those seeking a minor in Statistics. It was decided to emphasize the use of Statistics by those in other professions rather than direct the course to those doing research in Statistics. This is needed also, but it was decided to develop the module for the other professions first.

2.2 Sections of Course

Three general sections deal with (1) Encouraging students to realize that they are responsible for their own work and what comes out of it, (2) Encouraging students to evaluate what is their statistical level of competence (so that they will seek help from a statistician if necessary), and (3) Looking at the role of outside pressures in causing people to deviate from the ethical use of statistics.

2.2.1 Awareness of responsibility

Awareness of responsibility for one's research and publications therefrom is important to convey to the students. During their undergraduate days, students often relied on their parents or university professors to guide them in their decisions. Then in graduate school, they are more on their own and must make decisions which will have a lasting effect on their lives. Once articles are in print, they are a permanent record of what the student accomplished. The scientific community expects ethical behavior on their part both during their graduate days and throughout their professional careers.

2.2.2 Student evaluation of their statistical capabilities

The second major topic deals with the statistical capabilities of the students. They are asked to make an inventory of their statistical training and capabilities. Although many have taken a course or two in Statistics, they may not be ready to make decisions about the size of their experiment or sample survey or how to block effectively. In the module, they are given a number of questions, which might help them to decide whether or not they can go it alone without the help of a statistician. One set of questions might be:

1. Can you determine a proper sample size for your experiment?
2. Are you familiar with using contrasts for treatment comparisons?
3. Do you know the rules for dealing with outliers in a set of data?
4. Do you know when to use one-tailed tests and when to use two-tailed tests of significance?
5. Are you familiar with various methods of model selection in a multiple regression study?
6. Do you know when to use data transformations and which ones to use for particular situations?

If their answer to one or more of these questions is "No", then they are encouraged to seek the assistance of a statistician. A different set of questions could be developed for each application area.

2.2.3. *Pressures from outside which influence ethical behavior*

The third topic deals with the Pressures from the outside that might cause a researcher to bend and commit unethical acts. Often, the researcher would not commit these acts without outside influence. An example is in the form of expectations of a company paying for a research project to obtain results favorable to their company. Or, in a university setting, a new Assistant Professor might have considerable pressure to publish results which are worthy of acceptance by a journal and he or she may alter the results to make them more favorable to achieve journal acceptance of the paper. Master of Business Administration graduate students might succumb to pressures put on them to succeed rapidly and take short cuts putting aside ethical considerations. Related to the pressure issue is the question of what should be done when no positive results are obtained from a study. Of course full disclosure is the best policy but there is the question as to whether a journal will accept a paper describing such a study and the results therefrom. Some even might question whether the original hypothesis under examination in the study was a good one.

2.3 Ethics in Various Phases of Statistical Applications in Research

The module goes through all phases of the research process including ethical issues in the planning, analysis of data, interpretation and reporting.

2.3.1 Ethical issues in planning

The importance of planning is emphasized since without good planning, an experiment or survey may be doomed to failure from the start.

One case study involves the planning stage. A graduate student in Family and Consumer Sciences at a large university conducted a thesis research project to measure carpet wear as affected by such factors as pile material, cut vs. uncut pile, and pile height. She visited her statistician to obtain advice on the design of the experiment. The statistician developed a complicated experimental design which would control variation in carpet wear due to natural walking patterns, across the hallway where the experiment was located and which led to the student cafeteria. A diagram of the experimental plan is shown in Figure 1. The experimental design was a 3x2x2 factorial with factors: Pile Height, Fibers, and Finishes. The Finishes required a larger plot size so the experiment was set up as a Split-Plot design with Randomized Block arrangement of the whole plots. The sub-plots were then arranged in a 6x6 latin square arrangement (there were six blocks) in order to control position in the hallway. This was done because it was anticipated that most students would walk in the middle of the hallways. It should be noted that the cooperating statistician has since found some flaws in the randomization. The A and B finishes appear to be systematically arranged rather than at random. Also the right half of the design appears to be a mirror image of the left half.

The experiment went well except that at the end, the student tore down the experiment and failed

to record the position in the hallway where each carpet block was located. She had information on the treatment combination and the replication number, but lacked information on the position in the hallway (a non-treatment factor which was responsible for considerable variation in the experiment). It was possible to analyze the data using a much simpler model, but the advantages of the complicated model were lost. This meant less precision from the experiment.

The questions asked the students taking the statistics module of the ethics course were:

(1) Was the resultant failure of the experiment due to her independence or should the Statistician have spent more time with her explaining the design and how to mark the pieces of carpet regarding their position on the floor?
(2) Should her incorrect tear down of the experiment be grounds for denying her the Master's Degree?
(3) Would it have been better to use a simpler experimental design which the student could comprehend better?

2.3.2 Ethical issues in analysis of data

The module does not get into the fine points of which statistical analysis is appropriate for a given set of data or situation. In fact, alternative analysis options often exist. The module does emphasize that the researcher should learn to discriminate between situations in which he or she can make decisions about the analysis or those in which the assistance of a statistician is required. In many cases, the assistance of a statistician is required when the analyses are more complicated (and possibly more appropriate). Examples of the more complicated analyses include repeated measures, mixed models, spatial analyses and nonlinear models. There should be a clear understanding by both the researcher and the statistician as to which statistical analyses should be handled by which individual.

Three of the case studies in the module deal with the alleged omission of data that do not conform to the researcher's hypotheses. In all three, the results would have been adequately convincing without omitting data, but the researchers chose to make the case even more convincing. Two of them were high profile cases: Milliken's oil drop experiment and Gregor Mendel's pea experiment. Milliken won the Nobel prize in 1923 for his work in Physics and Mendel's original inheritance of traits distribution theory turned out to be correct. So the main result of the data omission in both case studies was a lowering of the variance. Students were asked to comment on the seriousness of the unethical omission of data in these three cases.

Another case study deals with the age-old question about accounting for every single person in the population in the censuses. In the early censuses, non-reponse was done on a 100% follow-up basis. However, in 2000, the Bureau of the Census followed up with only a sample on non-respondents. Some experts believe that the post census sample surveys degrade rather than improve the quality of the census data. There are political implications as the two major parties have different views about the importance of undercounting. The students are asked to analyze the implications of undercounting, especially with regard to whether it is a political problem or a statistical one.

Another case study dealt with a controversial environmental study conducted in the Skagerrak-Kattegat area in Scandinavia. Two highly respected, experienced marine biologists, both spending years in the field collecting and analyzing data from the Skagerrak accused each other of misusing Statistics. Dr. Alf Josefson and Dr. John Gray conducted separate long-term (6-17 years) studies of eutrophic levels in these bodies of water investigating changes in biomass in order to determine whether increased nutrient load from human activities were having an environmental impact. Josefson, collecting data from 14 different soft sediment stations below the halo cline in the Skagerrak, found a linear increase over time in 12 of them over time and extrapolated that, therefore, there was an overall increase in eutrophication in the Skagerrak area in general. Gray disagreed completely, saying that Josefson was not using his data properly. Although there were increases in biomass in 12 localities, only eight showed a significantly ($p<0.05$) positive increase and the increases in the other four were not significant. Gray argued against Josefson's results on two fronts. First, he noted that it is improper to pool data and extrapolate in this manner when some of the localities did not show significance. He indicated that it is irresponsible to publicize what he considered to be preliminary results of this sort. Josefson was guided by the precautionary principle -- publishing results from preliminary

studies when scientists feel that doing so will lessen the risks to either the environment or to the public. The case study raised some interesting questions about 1) the actual statistical analyses and 2) interpretation of the results, and 3) how and when to disseminate the information from research.

Students in the Ethics course were instructed that they didn't need to take sides in this controversy since there appears to be some vital missing information. For example, "how were the sample sites selected for monitoring and what population do they represent?" Josefson's original article did not make this clear. And why do all sampling sites need to show significant slopes in order to confirm eutrophication (this is implied by Gray's comments)? And why is it improper to pool data? We do this when combining results from a series of experiments.

2.3.3 Ethical issues in interpretation

Objectivity in interpretation is emphasized to the student taking the Research Ethics course. Emphasis should be placed on scientific conclusions and not mere speculations. Trustworthiness depends on objectivity. The problem is that the definition of objectivity varies by individual. As we have seen in the Skagerrak-Kattegat study, given the same analysis results, interpretations often vary by the person doing the interpretation.

Another case study involves both analysis and interpretation. A commercial firm made an arrangement with a university to conduct a sample survey for a contracted fee. Rather than having a senior faculty researcher take charge of the study, the leadership was given to a graduate student at the university. She used the occasion to put the company in a bad light by omitting key data from the analysis and then giving papers at various professional meetings in an attempt to show the company's product was inferior to another methodology. The company responded by asking for the original data so that they could reanalyze it and make a rebuttal. She was reluctant to provide the original data (which the company had paid for). University administrators responded to the company's inquiries that since the money the company paid for contractual work was not mixed with State-appropriated public funds, but rather was placed in the University's agricultural foundation, they had no responsibility to make the original data public. Students in the Ethics class were questioned on who owns the data in this case—the university, the graduate student or the company. They also were asked to comment on the ethics of omitting a key part of the data from the analysis to prove a point. Also, is this the sort of situation which would call for legal action by the company?

2.3.4 Ethical issues in the reporting phase

The Materials and Methods section in many publications is often very inadequate regarding details of the experimental design and statistical analyses methods used. In many cases, appropriate interpretation is not possible without such information. Also the reader cannot repeat the procedures in his or her own experiments if they are not adequately described. The statistical editor of the journal needs such a description to ascertain if proper statistical procedures have been used. Resnik (1998) and Whitback (1998) emphasize that honest disclosure is the key to integrity in Statistics.

One of the Senior Scholars developed an exercise for her Ethics class in Statistics in which students are given an article from a journal in the physical or biological sciences and asked to key out the Sources of Variation and Degrees of Freedom for an analysis of variance using only the information given in the Materials and Methods of the paper. In most cases, the students have found that the information given in the Materials and Methods is inadequate to key these out.

Often tables are missing measures of dispersion such as the standard deviation or standard error of a mean. Students are admonished to include measures of dispersion on the tables (or graphs) as they give the reader a knowledge of the precision achieved in the study. Misleading information can be conveyed by reporting excess decimals and by scaling graphs in such a way as to demonstrate the author's pet theory. Since original data are not often reported, the type of means reported is important. This is particularly relevant in reporting results from factorial experiments.

The topic of authorship is discussed in the module. Usually authorship is restricted to those who not only have participated in the research, but also have participated in the writing of the research paper. It is not ethical to include as co-

authors those who did not do any work in the study or writing. Credit can be given in the form of footnotes for such individuals. Statisticians can often fulfill their role by helping write the statistics section of the Materials and Methods and by carefully editing the portion of the paper showing results of statistical analyses. If they perform these duties, consideration should be given to including them as co-authors.

2.4 Poor Practice – Making a Mistake or Misconduct

Throughout the module, a distinction was made between deliberate fraudulent behavior and situations which arise from unintended mistakes, carelessness or lack of rigor. While the latter two may not fit the strict definition of misconduct, they are not examples of "good science" and they may have consequences which are similar to misconduct situations.

Bailar (1997) listed some practices which distort scientific inferences: failure to deal honesty with readers about non-random error (bias), inappropriate statistical tests and other statistical procedures, fragmentation of reports, low statistical power; suppressing, trimming, or "adjusting" data, undisclosed repetition of "unsatisfactory" experiments and, selective reporting of findings.

3. Summary

A one-hour credit course in Research Ethics designed to be taught to graduate students at Land Grant and other Universities was described. The development is a collaborative effort of eight Land Grant Universities. Although much of the course deals with general ethical principles, approximately one-seventh of the class time is devoted to participatory instruction in application modules, one of which is Ethical Use of Statistics. The course is directed toward majors in other fields who use Statistics in their research. Eventually, a separate module is planned for Statistics majors.

Three major sections of the module are (a) Encouraging students to realize that they are responsible for their own work and what comes of it, (2) Encouraging students to evaluate what is their level of competence and (3) Looking at the role of outside pressures in causing people to deviate from the ethical use of Statistics.

The approach used differs from the traditional Responsible Conduct of Research (RCR) approach involving compliance with a set of rules. Critical inquiry is nurtured and students are welcomed into the community of scholars and entrusted with a sense of community responsibility inherent in being a part of that community.

Ethical Use of Statistics is discussed in relation to all phases of the research process: planning, analysis, interpretation and reporting. Case studies are provided to illustrate principles and to elicit student discussion. Ethical standards vary according to application areas and this again dictates against encouraging students to memorize a set of rules and regulations for making decisions about conduct.

Distinction was made between fraud, misuse of Statistics and general sloppiness. Although all three might result in the same consequences, fraud is more readily connected with the term "unethical".

REFERENCES CITED AND READING MATERIALS

Bailar, John C. (1977), "Statistics and Deception" in Research Ethics: A Reader, (Deni Elliott and Judy E. Stein, eds. University Press of New England, Hanover).

Buhl-Mortensen and Stellan Welin (1998), The Ethics of Doing Policy Relevant Science: The Precautionary Principle and the Significance of Non-Significant Results, Science and Engineering Ethics, Opragen Publications, Surrey.

Cooper, B. Holmes (1990), The Honest Truth About Lying with Statistics. Charles C. Thomas, Springfield.

Deming. W.E. (1960), Sample Designs in Business Research. John Wiley & Sons, New York.

Huff, Darrel (1954), How to Lie with Statistics, W.W. Norton & Co. New York.

Leavitt, Fred (2001), Evaluating Scientific Research: Separating Fact from Fiction, Prentice Hall, Upper Saddle River.

Moore David (2006), Data Ethics An Introduction, From the Supplemental Material for David S. Moore and George P. McCabe: Introduction to the Practice of Statistics, 5th Ed. W.H. Freeman.

Pittinger, David (2001), Hypothesis Testing as a Moral Choice, Ethics and Behavior 11:2, 151-162

Resnick, David B. (1998), The Ethics of Science: An Introduction, Routledge, Oxford.

Whitbeck, Caroline (1998), Responsibility of Research Integrity, Chapter 6 in Ethics in Engineering Practice and Research. Cambridge University Press, Cambridge.

Figure 1.

Carpetware Study

Carpeting:
72 9" x 9" squares

Pile Height:
P_1 – High
P_2 – Low

3 Fibers:
M_1 – wool
M_2 – Acrilan
M_3 – Nylon

Finishes:
A – Uncut
B- Cut

Uncut
1A = $M_1 P_1$
2A = $M_1 P_2$
3A = $M_2 P_1$
4A = $M_2 P_2$
5A = $M_3 P_1$
6A = $M_3 P_2$

Cut
1B = $M_1 P_1$
2B = $M_1 P_2$
3B = $M_2 P_1$
4B = $M_2 P_2$
5B = $M_3 P_1$
6B = $M_3 P_2$