UNCERTAIN DECISIONS

Kelly Black
Clarkson College

1. INTRODUCTION

Statisticians tend to promote their field as a formal discipline to provide impartial insights, yet the field is more often identified as Benjamin Disraeli's third lie. The unfortunate truth is that despite the practitioner's lofty goal, the random nature of gathering information can result in errors that can lead even the most careful researchers astray. An important question that must be addressed is how do researchers explain this to their intended audience.

The answer is not simple and has conflicting priorities. Researchers must let their audience know the possible deficiencies of their methods but should not do so in a way that can shroud the potential issues with a blizzard of caveats and warnings. Professional societies have explicit guidelines to aid researchers in discussing their results, but the guidelines are problematic when readers have a wide range of sophistication in their understanding of the statistical methods.

At the same time the full burden of communication does not fall on those who convey information. It is important that the consumer of statistical information have a basic understanding of statistical techniques. This was made explicit by Douglass Curran-Everett, Sue Taylor and Karen Kafadar who addressed the American Physiological Society with these words:

> If we fail to understand fully these fundamental statistical concepts—if our statisticalreasoning is faulty—then we are more likely to reach wrong scientific conclusions. Wrong conclusions based on faulty reasoning is shoddy science; it is also unethical (Curran-Everett et al., 1998).

There is a distinct advantage to discussing ethics in the context of an introductory statistics class. People make important decisions based on statistics and often turn to statistical methodologies to gain insight into

how to make difficult choices. For example, at the time I was writing this article, I opened *The New York Times* to find an article about administrators struggling with ways to balance the distribution of medical supplies in the United Kingdom (Harris, 3 December 2008).

The teaching of ethics in the context of statistics is important with respect to another aspect. It is extremely easy to gather data, and data are everywhere. Unfortunately, gathering data correctly and performing appropriate, careful analyses can be extremely difficult. It is important for citizens to understand the nature of data as well as the limitations and pitfalls of statistical methods.

In light of the importance of sharing the results of statistical analyses it is important to note that a continuing trend in the statistics community is to examine how to discuss statistical results (Petocz and Reid, 2005). In this paper we share some of our efforts to introduce an ethics component into our introductory statistics course. We first discuss the nature of statistics and what is the primary goal of the introductory statistics course. Next, we discuss the issues in sharing and communicating the results of statistical analyses followed by a discussion of the trade-offs that must be considered when discussing statistical results. Finally, we offer a description of the history of our efforts to bring these things together in our introductory statistics course.

2. What Is Statistics?

One of the primary goals of a statistician is to provide a set of coherent and reliable tools to aid in making decisions. The nature of the tools requires a tight integration of experiment and analysis. Careful planning is required prior to the execution of the experiment, and careful evaluation is required during and after the experiment. One does not collect data and then perform a set of analyses on the harvested evidence. The methods used to gather information have an impact on the subsequent analysis, and the researcher must consider this aspect prior to performing an experiment.

It would overwhelm the students to take part in the full integration of analysis and experiment in one introductory course. Instead, we set modest goals for the introductory statistics course. The focus is to obtain a basic understanding of the nature of data and to provide an overview of the basic ways to describe and analyze data. The students are exposed to the most basic ways in which statistics are used in the decision making

process. In particular the students are exposed to the ideas of the "margin of error" and "statistical inference."

One of the primary difficulties in trying to share the basic ideas in the course is that students enter the course with a number of firm preconceptions about statistics and about the meaning of the underlying ideas. Often these ideas are incorrect. For example, one of the most basic preconceived notions that our students have about statistics is that the subject is about "numbers." The paradox is that it is easy to confront the students' misconceptions but difficult to change them.

One of the students' most cherished misunderstandings is that calculations based on sampled "data" represent a clear picture of the world around us. The very act of sampling, however, introduces randomness into what we see. Put simply, taking a sample is akin to rolling the die, and results from the analysis of sampled data do not provide a definitive picture of the underlying physical processes.

One way in which we confront the idea of sampling proportion is by looking at results for a simple sampling experiment such as flipping a coin or choosing beans of different colors from a jar. In the context of ethics we examine case studies that have conflicting results. For example, in one case study we look at published results on the incident rate of parotid gland tumors and cell phone use. We explore the results from two papers. In the first paper, (Lonn *et al.*, 2006) the researchers' sample indicates that cell phone use does not result in a higher rate of parotid gland tumors. In the second paper (Sadetzki *et al.*, 2008) the sample indicates that cell phone use does result in a higher rate of parotid gland tumors. Put in the context of public health the idea of interpreting a sample proportion impacts people's lives and is an ethical issue.

3. Sharing the Results of Statistical Analyses

The random nature of sampled data implies that it is possible to do everything correctly and still reach the wrong conclusions. The implication is that it is not possible for researchers to make a definitive statement based on statistical analysis and not convey the whole truth to their readers. When trying to provide readers with the necessary information to make informed decisions, researchers must make it clear that there is some probability that the calculations that are made do not reflect the actual response to a given set of treatments.

One of the factors that compounds the difficulty faced in sharing statistical results is that different fields of specialization have different

cultures and different expectations on how results are discussed. One of the common features across different disciplines is that there is an expectation that a clear and concise explanation of results be provided, but just giving the numbers can be misleading (Fischhoff, 1995).

A simple presentation of the results does not provide a complete picture. At the same time the researcher must find a balance between the desire for clarity and the desire to inform. As an example of how difficult this balance can be one need only look at the list of warnings and caveats that comes with any medication.

An additional factor that comes into play is that researchers can do everything right and still reach the wrong conclusion through no fault of their own other than sheer "bad luck." What makes this aspect most insidious is that the researchers may not know in what way they are wrong. In the parlance of statistical inference, tools such as the "power" of a test and the "*p*-value" are used to help researchers understand what could be happening. Neither of these tools provides the actual probability that the conclusions are correct; rather they provide conditional probabilities.

Fortunately professional standards exist to aid researchers.[1] The standards can vary from discipline to discipline, but they provide a framework to help guide researchers. One of the basic principles that is common in the different guidelines is that we are expected to provide enough redundant information to allow others to confirm key results. It is important to keep in mind that what constitutes enough redundant information is not always obvious, so the researchers must focus on the nature of their results and the subsequent presentation.

A great deal of the literature in a given field tends to include a relatively narrow set of statistical analyses and practices. Despite the use of common techniques it is still necessary for researchers to step back and think about what kind of information they are sharing. There are a number of potential pitfalls waiting for researchers trying to share their results.

One potential issue is that the methodology or techniques used in a given study are inappropriate (Hotz, 14 September 2007). The reason that researchers are expected to disclose so many details is that they must allow others to determine the level of trust that for a given set of results. Given the volume of problematic results that are published the readers of such materials must have a certain basic understanding of the statistical analyses and perform some minimal level of due diligence when examining a given set of results.

One example that we examine in class is a number of articles on the cholesterol levels of American adults (Tanner, 13 October 2005; Peggy Peck, October 11, 2005; Warner, 12 October 2005). The articles indicate that the use of a new class of drugs has resulted in a decline in LDL cholesterol levels. In the classroom we go back and examine the original paper (Margaret D. Carroll *et al.*, 2005). The paper is part of a series of papers exploring cholesterol levels over time. The authors report a decline in LDL cholesterol levels for some age groups and suggests that the use of a new class of drugs might contribute to the decline. We look further at the results reported in the original paper and examine the trend over a longer time period. It soon becomes evident that the decline is part of a broader trend, and the subsequent focus on one particular cause is a gross simplification of the bigger picture that is emerging.

Maintaining a critical view of a set of results can be a difficult task. It requires a high level of understanding of the statistical methods and requires a great deal of attention to detail. For example, subtleties such as the difference between odds ratios and risk ratios are important (Cummings and Rivara, 2003). An additional issue for researchers is that our understanding of the statistical methods changes over time. Consequently the ways in which ideas and results are communicated also change over time (Altman, 2000). Not only should practitioners in the field have a basic understanding of the statistics, but they should be aware of trends and changes in the field. An explicit discussion of these issues in the context of ethics provides direct support for why we consider the different aspects of the statistics curriculum.

4. How Much Information Is Enough?

Sharing statistical results requires a delicate balance. Providing too little information can mislead the reader. Providing too much information can obfuscate and hide important information. Fortunately there are some basic guidelines and standard practices that can help researchers decide how much information to provide.

One of the goals in the classroom is to first share these basic guidelines and to make students aware that there are cultural differences between disciplines. Placing the discussion in the context of ethics provides motivation for why there is a need for such practices. More importantly, it provides motivation as to why students should think deeply about current practices and why students should be aware of how practices are changing.

Our ethics context focuses on some relatively basic questions. The primary issue is about misleading readers. This is a non-trivial issue to address and requires a discussion about the difference between misleading and lying. This discussion is made more difficult in our context because it is not easy to find guidance from the ethics communities on how to address this aspect in the presence of uncertainty.

Ultimately we focus on a few basic goals. The first goal is to be clear and concise. It is far too easy to obfuscate a conclusion using the technical language of statistics. An additional goal is to make sure the readers are informed of the underlying assumptions. The majority of statistical techniques rely on some assumptions about the underlying data, and it is important to identify the assumptions and make a judgement as to how sensitive and how robust a technique is with respect to those assumptions. Finally, it is important to provide just enough redundant information to allow readers to reproduce a given set of results. This allows readers to check if certain calculations are correct, and more importantly it allows readers to make their own calculations that may be more appropriate for their needs.

4.1 Lying With a Graph

There are many different ways to share statistical results. The primary techniques are graphical and analytical. The presentation of graphical results is a special case. The ease in which graphical representations can be generated makes it more important than ever to ask fundamental questions about any graph that is presented.

It is extremely easy to mislead using a graph. The human eye responds to area and length scales in different ways. For example, the choice of a pie chart versus an histogram can convey very different meanings (Huff and Geis, 1993). Simple changes like altering the scale or shifting the scale can also convey radically different meaning.

Also, the simple act of adding a little extra bit of decoration on a graph can make it harder to read or mask the information presented (Tufte, 1992). The practice of trying to "punch up" a graph can change how the information is interpreted. There are also important technical details to consider when making a graph. For example, Tufte has developed metrics that can be used to compare how a particular graph can skew the information conveyed (Tufte, 1992).

4.2 Misleading With Quantitative Results

In addition to graphical techniques for displaying the characteristics of data, there are also numerous quantitative measures used to describe different aspects of data. Many of these measures are interrelated, and it is important that enough information be given to the reader to provide a broad understanding of the data. For example, just providing a sample mean does not convey meaningful information without an conveying an understanding of the spread of the data.

In many different fields a number of expectations have evolved in terms of what information is provided in the literature. In these cases the guiding principle is that enough redundant information be given so that readers can verify and perform their own calculations. In the context of an ethics discussion, one question that arises is whether or not simply following the guidelines is an adequate way to inform the readers. For example, in most science fields it is common practice to provide the *p*-values that result from a given analysis.

Our class discussion on this topic revolves around whether or not providing *p*-values is a meaningful measurement. The first issue is that despite its wide-spread use there is evidence that it is not a well understood quantity (Goodman, 1999). The fact that the *p*-value is a conditional probability may not be fully understood by many researchers.

The second issue is that a number of researchers have raised questions about the reporting of *p*-values (Sterne and Smith, 2001). The questions center on what information the *p*-value tells us as well as the logical fallacy in their use of understanding the level of an impact as opposed to simply determining statistical inference (Cohen, 1994). The majority of statistical textbooks discuss the use of *p*-values as a tool to understand the basic issues associated with sampling and inference, but in a classroom discussion on ethics this is an opportunity to explore how common practices continue to evolve in time.

As an example, the ethics discussion provides an opportunity to explore current practices and read a paper on how active researchers themselves communicate their own views. One paper that we explicitly explore is the paper by Nakagawa and Cuthill (2007). The ethics context allows us to ask fundamental questions about the meaning of the analysis that is given in the textbook. In this context it is no longer "just another technique."

Another issue related to the use of quantitative measurements of data is to ask basic questions about the techniques that are used. One example that we use in our classroom discussion is the impact of heaping

versus rounding on the resulting quantitative results (Crockett and Crockett, 2006). In the context of an ethics discussion, this is an opportunity to discuss the basic goal of insuring that the readers understand the full scope of what operations we have done on the data.

The ethics discussion has a direct relationship to the mathematics that the students are expected to understand. These issues of the techniques have a direct bearing on how robust a particular method is. The "robustness" of a particular method is a question of how sensitive a given technique is to violations of the underlying assumptions about the data. The explorations of the ethical issues lead directly to important aspects of the statistical methods and provide concrete motivation as to why such issues are important.

5. Our Experience

We are currently beginning the third iteration of the incorporation of ethics into the introductory statistics class. In our first iteration, we tried to do it in a way that was consistent with the textbook. Many books have an early section or chapter in which the various issues are discussed. Our goal is to motivate the development of the subsequent techniques and analyses. At the same time we wish to maintain a level of consistency with the other activities. The augmented ethics component was introduced in a way that was consistent with the book, and we took part in a focused classroom discussion on the relevant ethics issues.

The results were not satisfactory. The ethics component became part of the course checklist for the students, and it was not properly integrated with the relevant material. The ethics materials were not seen as relevant to the rest of the course and were not placed in their proper context. Worst of all, the ethics component did not provide a context to more fully explore and motivate the mathematics.

In the second year, we have broken away from the treatment in our textbook. The relevant materials are covered later in the term, and the ethical issues are discussed in the classroom as they arise in the context of the statistical exploration. In this way the ethical issues are intermingled with the mathematical considerations. Most importantly, the ethical issues offer a way to reinforce and motivate the mathematical considerations. Ethics provides a gateway to a more meaningful discussion of the mathematical issues and leads to a deeper exploration of the statistical issues.

6. Conclusions

Ethical issues are an important aspect of the statistics curriculum. The practices that have evolved are the direct result of concerns of the ethical dimension of how statistics is used in practice. The important ethical issues are often included in the statistics course, but the issues are not always fully explored or fully integrated into the curriculum.

We have included ethical aspects to a much greater degree in our introductory statistics course. We do so by integrating the ideas throughout the curriculum, and use ethical considerations as a way to examine statistical techniques in a different context. More importantly we use the ethical considerations as a gateway to explore fundamental mathematical concepts.

Our focus is on the ways in which the results of statistical analyses are shared. By focusing on the presentation of statistical results we bring current practices into sharp focus which also brings into question why such practices have evolved and why they continue to change. Such considerations offer an invaluable context that provides a framework for the whole course and reinforces the whole of our activities.[2]

Notes

[1] An excellent example is the ASA Ethical Guidelines, *http://www.theasa.org/ethics.htm.*

[2] Special thanks go to Dr. Anastasia Pease, the director of the Ethics across the Curriculum program at Union College. She provided immense help in the initial literature search as I began the task to integrate the ethics component into the statistics course. She also provided invaluable insight and support in the writing of this document. Additional thanks go to Michael S. Rapaport whose support made it possible to expand the efforts to include ethics in the curriculum across Union College.

References

Altman, D. G., 2000, Statistics in Medicine **19**(23), 3275.
Cohen, J., 1994, American Psychologist **49**(12), 997.
Crockett, A., and R. Crockett, 2006, Historical Methods **39**(1), 24.

Cummings, P., and F. P. Rivara, 2003, Arch Pediatr Adolesc Med **157**(4), 321, *http://archpedi.ama-assn.org/cgi/reprint/157/4/321.pdf.*

Curran-Everett, D., S. Taylor, and K. Kafadar, 1998, Journal of Applied Physiology **85**(3), 775.

Fischhoff, B., 1995, Risk Analysis **15**(2), 137.

Goodman, S. N., 1999, Annals of Internal Medicine **130**(12), 995.

Harris, G., 3 December 2008, British balance gain against the cost of the latest drugs, *http://www.nytimes.com/2008/12/03/health/03nice.html ?em*#, last accessed December 3, 2008.

Hotz, R. L., 14 September 2007, Most science studies appear to be tainted by sloppy analysis, *http://online.wsj.com/article/SB118972683557627104.html*, last accessed December 2, 2008.

Huff, D., and I. Geis, 1993, *How to Lie With Statistics* (W. W. Norton & Company).

Lonn, S., A. Ahlbom, and H. C. Christensen, 2006, American Journal of Epidemiology **164**(7),637.

Margaret D. Carroll, M., M. David A. Lacher, P. Paul D. Sorlie, M. James I. Cleeman, P. David J. Gordon, MD, M. Michael Wolz, P. Scott M. Grundy, MD, and M. Clifford L. Johnson, 2005, The Journal of the American Medical Association **294**, 1773.

Nakagawa, S., and I. C. Cuthill, 2007, Biological Reviews **82**(4), 591 .

Peggy Peck, M. T., Senior Editor, October 11, 2005, Cholesterol is hitting new lows for americans, *http://www.medpagetoday.com/Cardiology/Dyslipidemia/1915*, last accessed January 29, 2009.

Petocz, P., and A. Reid, 2005, in *Statistics Education and the Communication of Statistics* (International Association for Statistical Education, Sydney, Australia), *http://www.stat.auckland.ac.nz/~iase/publications/14/petocz.pdf.*

Sadetzki, S., A. Chetrit, A. Jarus-Hakak, E. Cardis, Y. Deutch, S. Duvdevani, A. Zultan, I. Novikov, L. Freedman, and M. Wolf, 2008, American Journal of Epidemiology , **457**.

Sterne, J. A. C., and G. D. Smith, 2001, Physical Therapy **81**(8), 1464.

Tanner, L., 13 October 2005, Statin drugs reason for fall in u.s. cholesterol, *http://www.redorbit.com/news/health/271044/statin_drugs_reason_for_fall_in_us_cholesterol/index.html*, last accessed January 29, 2009.

Tufte, E. R., 1992, *The Visual Display of Quantitative Information* (Graphics Press), second edition.

Warner, J., October 12, 2005, Americans' cholesterol levels declining, *http://www.foxnews.com/story/0,2933,172026,00.html*, last accessed January 29, 2009.