THERE'S MORE TO STATISTICS THAN COMPUTATION—TEACHING STUDENTS HOW TO COMMUNICATE STATISTICAL RESULTS

PECK, Roxy
California Polytechnic State University,
USA

Statistics is a discipline whose primary purpose is to provide analysts and researchers with methods for collecting data in a reasonable way and for making sense out of data. As such, the role of context is critical and the importance of context is among the things that distinguish statistics from mathematics. The importance of context is a primary reason that communication is such an important aspect of statistics problems. In statistics, meaning comes from context, and it is the interpretation of the analysis in context that is the ultimate desired outcome of analyzing data. Even though communication is an important part of interpreting and analyzing data, traditional statistics courses often focus on the computational aspects of data analysis and often neglect issues of effective communication. This paper considers ways to develop "communication basics" in the introductory statistics curriculum.

INTRODUCTION

Statistics is a discipline whose primary purpose is to provide analysts and researchers with methods for collecting data in a reasonable way and for making sense out of data. As such, the role of context is critical and the importance of context is among the things that distinguish statistics from mathematics. Although data is often numerical, George Cobb and David Moore (1997) remind us that "data are not just numbers, they are numbers with a context", and go on to say

> Although mathematicians often rely on applied context both for motivation and as a source of problems for research, the ultimate focus in mathematical thinking is on abstract patterns: the context is part of the irrelevant detail that must be boiled off over the flame of abstraction in order to reveal the previously hidden crystal of pure structure. *In mathematics, context obscures structure.* Like mathematicians, data analysts also look for patterns, but ultimately, in data analysis, whether the patterns have meaning, and whether they have any value depends on how the threads of those patterns interweave with the complementary threads of the story line. *In data analysis, context provides meaning.*

It is the importance of context that makes communication is such an important aspect of statistics problems. In statistics, meaning comes from context, and it is the interpretation of the analysis *in context* that is the ultimate desired outcome of analyzing data. It is the interpretation of results and the communication of the results in context that require good “statistical communication” skills.

Even though communication is an important part of interpreting and analyzing data, traditional statistics courses often focus on the computational aspects of data analysis and neglect issues of effective communication. Graphing calculators and statistical software are now capable of performing sophisticated calculations with little effort required on the part of the calculator or computer operator. These days, if all a student can bring to the table is the ability to perform the computational aspects of data analysis, they can be easily replaced by an inexpensive calculator. To actually make an intellectual contribution to the data analysis process, they must be able to understand the importance of appropriate planning in the data collection, and the relationship between the type of data collected, the data collection methods, and the appropriate analysis tools. Moreover, they must be able to draw meaningful conclusions that connect context and the analysis, and communicate those results to others.

In addition to the fact that interpretation and communication are integral to data analysis, there is another reason for students to learn to effectively communicate statistical ideas and results, namely learning to explain statistical analyses and concepts in different settings and to different audiences which will further enhance the student's own conceptual understanding of statistics.

There is also a growing trend, in the U.S. and elsewhere, to include interpretation and communication as part of student assessment in introductory statistics courses. A prominent example is the rubric used to score free response questions on the Advanced Placement Statistics exam in the U.S. Advanced Placement (AP) Statistics is a year-long course taught using a common course outline provided by The College Board. It offers capable secondary students the opportunity to take a college level course in introductory statistics while still in high school. Students who complete this course and who pass a national exam can obtain college credit and/or advanced placement in statistics at most colleges and universities in the U.S. In the scoring of AP Statistics exam (taken by nearly 70,000 students in 2004), both statistical knowledge and communication are considered in evaluating student solutions to the free response questions. The scoring guide for free response questions that appears in the Teachers Guide for AP Statistics (2004) identifies and describes five different levels of achievement for statistical knowledge and for communication. This document also includes a section titled "The Communication Dimension of the Free-Response Questions", to assist teachers in recognizing the importance of emphasizing statistical communication when teaching this course.

If interpretation and clear and efficient communication are essential aspects of data analysis, it makes sense to look critically at introductory statistics course to assess where these skills might be addressed and to consider ways in which statistical communication skills might be more effectively developed in students. This paper offers some suggestions for teachers of introductory statistics courses.

DEVELOPING STATISTICAL COMMUNICATION SKILLS

Effective communication of statistical analyses and concepts requires thought, judgment, and most of all, experience. Most students who are just learning statistics don't have much experience with data and have not yet developed a systematic way of thinking about data.

While most teachers of statistics understand the importance of modeling interpretation and effective communication in classroom examples and a good textbook will also consistently model interpretation and communication in its data analysis examples, students do not always find this type of "modeling good practice" to be enough to enable them to identify the characteristics of good communication and then to reproduce it themselves. Students often try to come up with generic, "cookie-cutter" statements that they will then try to apply regardless of context. And, they often don't see the need to go beyond making statements like

> The mean and standard deviation of (*put your variable name here*) were *XXX* and *XXX*, respectively.

or

> There is sufficient evidence to reject the null hypothesis that the mean (*put your variable name here*) is *XXX.*

So, what can we do to get students to move beyond making statements like those above and to reflect on the meaning for the context of interest? Some suggestions follow.

1. Be explicit about what is needed for good communication in different settings. For example, the AP Statistics Teacher's Guide referenced earlier offers suggestions as to what elements should be present when describing distributions, describing a study, etc. Some texts, such as Devore and Peck (2005) and Peck, Olsen and Devore (2005), now include sections specifically on communication of results of statistical analyses

to help students understand what needs to be addressed when describing a study design, how to communicate effectively with graphs and what kinds of things should be included in a verbal description of the data distribution displayed, etc. For example, the following paragraph is taken from the chapters on graphical displays in the two texts previously referenced.

When reporting the results of a data analysis, a good place to start is with a graphical display of the data. A well constructed graphical display is often the best way to highlight the essential characteristics of the data distribution, such as shape and spread for numerical data sets or the nature of the relationship between the two variables in a bivariate numerical data set.

For effective communication with graphical displays, some things to remember are:

- Be sure to select a display that is appropriate for the given type of data.
- Be sure to include scales and labels on the axes of graphical displays.
- In comparative plots, be sure to include labels or a legend so that it is clear which parts of the display correspond to which samples or groups in the data set.
- While it is sometimes a good idea to have axes that don't cross at (0, 0) in a scatter plot, the vertical axis in a bar chart or a histogram should always start at zero (see the cautions and limitations later in this section for more about this).
- Keep your graphs simple. A simple graphical display is much more effective than one that has a lot of extra "junk". Most people will not spend a great deal of time studying a graphical display, so its message should be clear and straightforward.
- Keep your graphical displays honest. People tend to look quickly at graphical displays and so it is important that a graph's first impression is an accurate and honest portrayal of the data distribution. In addition to the graphical display itself, data analysis reports usually include a brief discussion of the features of the data distribution based on the graphical display.
- For categorical data, the discussion might be a few sentences on the relative proportion for each category, possibly pointing out categories that were either very common or very rare when compared to other categories.
- For numerical data sets, the discussion of the graphical display usually summarizes the information the display provides on three characteristics of the data distribution—center or location, spread, and shape.
- For bivariate numerical data, the discussion of the scatter plot would typically focus on the nature of the relationship between the two variables used to construct the plot.
- For data collected over time, any trends or patterns in the time series plot should be described.

While many of the “things to remember” will seem obvious to teachers of introductory statistics, they are not always obvious to students after seeing just a few examples presented in class and in the textbook. If we can be explicit about what constitutes good practice in statistical communication is various settings, either in class or in the assigned reading, students will have an easier time developing an organized way of thinking about data.

2. Emphasize the importance of context. Insist that students communicate in context and that they go beyond the cookie cutter interpretations to indicate what the results mean in the context of the problem. Use data (numbers in context) not just numbers for all examples and assignments incorporated into the course. Model good communication for your students. When doing examples in class, don’t stop when the mechanics are finished. Make sure to always carry through to a conclusion or interpretation in context.

3. Ask questions that require explanation and interpretation throughout the course—not just in inference. Interpretation and communication are just as important in early course topics as they are toward the end when inferential methods are covered.

4. Don't accept "mechanics only" answers as correct on homework or exams. If you want students to understand that communication is important, it is important that it be a part of the assessment process. Don't accept calculator or computer package "talk". "I entered the data into my calculator and then did a one-prop z test" is not a good explanation of the analysis!

5. Encourage students to read as well as to write. Reading the textbook helps students see how to communicate statistical ideas.

6. Ask students to write about statistical processes—for example, ask them to write a paragraph explaining how you tell if there are outliers in a data set, or how they would carry out a simulation to estimate a probability, or why random assignment is important in the design of an experiment.

Good statistical communication, like so many other things in life, improves with practice. If you can help students understand why effective communication is essential to data analysis and to their ability to make a meaningful intellectual contribution to the data analysis process and if you give them frequent opportunities to practice, you may be pleasantly surprised by what they can do by the end of the course. And, while it may be painful to read students' early attempts, practice, consistent expectations and good feedback can produce rewarding results.

REFERENCES

Cobb, G.W., & Moore, D.S. (1997). Mathematics, statistics, and teaching, *American Mathematics Monthly*, *104*, 801-823.

Devore, J., & Peck, R. (2005). *Statistics: The exploration and analysis of data* (5th edn.). Duxbury Press, Belmont, CA.

Olsen, C. (2004). *Teacher's guide for advanced placement statistics*. The College Board, NY.

Peck, R., Olsen, C., & Devore, J. (2005). *Introduction to statistics and data analysis* (2nd edn.). Duxbury Press, Belmont, CA.

Traditionally, statistical computing courses have taught the syntax of a particular programming language or specic statistical computation methods. Since the publi-cation of Nolan and Temple Lang [2010], we have seen a greater emphasis on data manipulation, reproducible research, and visualization. This shift better prepares students for careers working with complex datasets and producing analyses for multiple audiences. But, we argue, statisticians are now often called upon to develop statistical software, not just analyses, such as R packages implementing new analysis methods or machine lear Slides and videos for Statistical Learning MOOC by Hastie and Tibshirani available separately here. Slides and video tutorials related to this book by Abass Al Sharif can be downloaded here. "An Introduction to Statistical Learning (ISL)" by James, Witten, Hastie and Tibshirani is the "how to" manual for statistical learning. Inspired by "The Elements of Statistical Learning" (Hastie, Tibshirani and Friedman), this book provides clear and intuitive guidance on how to implement cutting edge statistical and machine learning methods. ISL makes modern meth How Student Learn Statistic - Free download as PDF File (.pdf), Text File (.txt) or read online for free. How Student Learn Statistic.Â Regardless of the setting, a major concern of those who teach statistics is how to ensure that the students understand statistical ideas and are able to apply what they learn to real-world situations. Although teachers of statistics often express frustration about difficulties students have learning and applying course material, many may be unaware of the growing body of research related to teaching and learning statistics. In this paper I attempt to summarize this literature and apply it specifically to improving learning outcomes in college-level statistics courses.