1-30-2015

Emphasizing the Entire Research Process Throughout the Curriculum: The Next Step in Real Data Integration in Introductory Statistics Courses

Nathan L. Tintle
Dordt College, nathan.tintle@dordt.edu

Recommended Citation

Tintle, N. L. (2015). Emphasizing the Entire Research Process Throughout the Curriculum: The Next Step in Real Data Integration in Introductory Statistics Courses. Retrieved from https://digitalcollections.dordt.edu/faculty_work/816

Emphasizing the Entire Research Process Throughout the Curriculum: The Next Step in Real Data Integration in Introductory Statistics Courses

Abstract

Access blog post at publisher's site:

https://www.causeweb.org/sbi/?p=617

Keywords

simulation-based curricula, inference, college students, courses of study

Disciplines

Statistics and Probability

Emphasizing the entire research process throughout the curriculum: The next step in real data integration in introductory statistics courses

In 2005, the [Guidelines for Assessment and Instruction in Statistics Education](#) made six recommendations about how we should teach introductory statistics. One of these recommendations is to use real data. The report goes on to argue that real data, as opposed to merely realistic (made-up for a hypothetical context) or naked (no context provided) data, is preferred. I argue that we should go a step further by emphasizing the entire statistical research process throughout the curriculum.

To ensure our students leave our courses recognizing the indispensable nature of statistics in science and society we must force ourselves to get out of the box and embrace teaching the entire research context by utilizing real data that matters.

Moving beyond realistic data

The practice of providing data without context is common in courses focused on mathematical thinking (e.g., deductive reasoning) as opposed to statistical thinking (e.g., inductive reasoning incorporating chance and variability). For example, if a learning objective of a course is to have students utilize equations to compute the estimated slope and intercept of a least-squares regression line given a dataset of paired x and y values, the context of the data is not really relevant. Realistic data, where data is made up but a hypothetical context is placed on the data, serves as an incremental step towards statistical thinking. However, the use of real data is even better.

Using real data in introductory statistics courses makes it clear to students that statistics is not an abstract, fanciful discipline that has no bearing on the real world, while simultaneously opening the door up to potentially interesting discussions about the 'bigger picture' of statistics (impact of data production, what the researchers should do next, why the researchers did the study in the first place). Unfortunately, simply using real data does not go far enough. I once had a top bio-medical researcher tell me that when she first took introductory statistics she spent much of her time in the course talking about M&M colors, and it wasn't until graduate school that she really appreciated how vital statistics was to the entire scientific research community. There is no question that having students do cute in-class activities can be engaging and fun for all, but if 'real data' is too often disconnected from real research, we run the risk of leaving our students seeing little connection between their primary disciplines and statistical concepts, and missing a major opportunity to teach statistical thinking.

but if 'real data' is too often disconnected from real research, we run the risk of leaving our students seeing little connection between their primary disciplines and statistical concepts, and missing a major opportunity to teach statistical thinking.

How do we get beyond 'real, but cute' and emphasize statistical thinking to students? To do this, I suggest that we should move from simply 'real data' to 'real data *that matters*," while making sure that we highlight the entire statistical investigation process with nearly every context we present to students.

Although there is a place for limited use of cute data in introductory statistics courses, there are ways to do in-class data collection activities that are both engaging, and have a broader and more important data context. For example, in the *ISI* curriculum we do an activity where students look at pictures of two different male faces. Students are asked to associate one of two names with the faces (Bob or Tim; see the full activity in Chapter 1 posted here). Class data is compiled to see how many students associated a particular name with a particular face. This in-class data collection is modelled after a real peer-reviewed journal article, and while 'cute,' also emphasizes to students that this is real research (published here) and opens the door up to class discussions about Psychonomics (the psychology of names) and its broad social implications.

So, why don't we do activities like this more in our classes? I think that there are two main reasons. First, it's hard. To find real research studies that are simultaneously engaging and accessible takes time. It is unrealistic to think that individual instructors should or can be in the business of combing the applied research literature to find interesting examples. Fortunately, there are more and more textbooks using real research studies and many other web-based resources (e.g., repositories like CAUSEweb or listservs like isostat) where good datasets and examples are shared regularly.

Second, teaching this way puts many instructors outside of their comfort zone by opening up the door to interdisciplinary, contextual discussions at the interface of statistics and other disciplines. Statistics instructors wonder how they can be expected to lead discussions involving another discipline (e.g., Psychonomics) and/or that too much class time will be taken away from the core 'statistical' objectives of the course. The good news is that you don't have to be an expert. Simply applying good statistical thinking to research studies (e.g., What motivated the researchers to do this study? What are the implications of the study design? What do you think the researchers will do next?") puts instructors at least one or two steps ahead of most students when it comes to discussing real research in the classroom. Furthermore, we should not underestimate the pedagogical value of empowering students who have potentially more disciplinary expertise than we do when discussing a particular study. I have had many times in my classes where a weaker student's face lights up and they offer stunningly deep and important contextual insights when discussing a real research study because of their expertise in the context being discussed.

I have had many times in my classes where a weaker student's face lights up and he or she offers stunningly deep and important contextual insights when discussing a real research study because of previous expertise in the context being discussed.

For 'math phobic' students this approach goes a long way to helping students feel at home, welcome, and confident in our classrooms. Finally, in addressing the critique that talking about real research takes too much time away from the main objectives of the course, we must remember that statistical thinking should be a core objective of our courses and this thinking is taught, perhaps best, through data story-telling, not discrete, disconnected, mathematically-focused, learning objectives. The time is well worth it. There are other things to cut from our curricula before this.

Suggestions

Let me conclude by offering two key ways to make 'real data that matters' discussions a part of your class. First, motivate new course concepts with real research studies and highlight the entire statistical investigation process with (nearly) every study you talk about. In our introductory statistics curriculum, the first topic that students encounter is the six-step statistical investigation process, a process that mirrors the scientific method, implicitly demonstrating to students that statistics is an integral part of scientific reasoning from the beginning (click to enlarge):
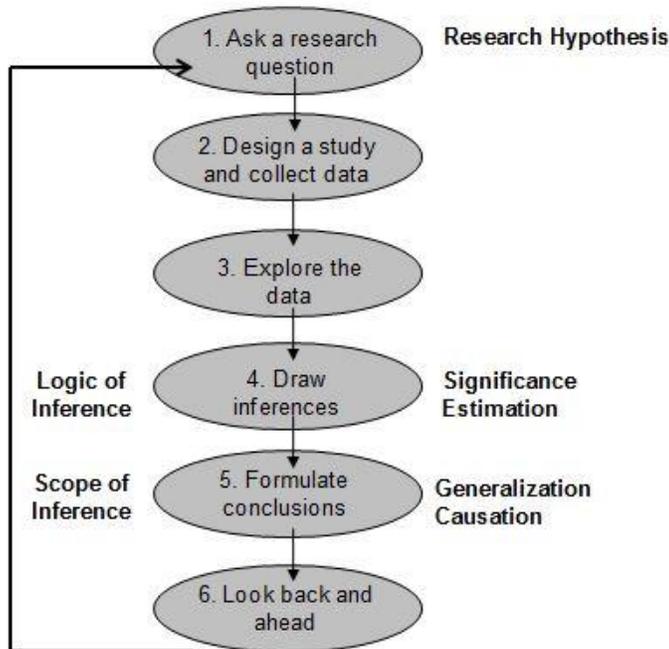
With nearly every study we examine, we pointedly ask or show students all six steps, even though our key emphasis may only be on one of the steps in any particular section of the book. Second, consider the use of simulation-based inference methods to facilitate more 'real data that matters' discussion. In a traditionally organized curriculum, emphasizing the entire statistical process can be hard because inference is not introduced until late in the course. Courses using simulation-based inference have the ability to more quickly move students to inference questions, and, thus, classes using simulation-based inference can have a deeper and richer discussion about the entire statistical investigation process earlier in the course. For a more

fleshed out discussion of this final point in the context of the entire undergraduate statistics curriculum read our recent white paper [here](#).

To ensure our students leave our courses recognizing the indispensable nature of statistics in science and society we must force ourselves to get out of the box and embrace teaching the entire research context by utilizing real data that matters.