Some Key Comparisons between Statistics and Mathematics

and Why Teachers Should Care

Allan Rossman¹, Beth Chance¹, Elsa Medina²

¹Department of Statistics, ²Department of Mathematics

Cal Poly – San Luis Obispo
Statistics is a mathematical science. That sentence is likely to be the shortest in this entire article, but we want to draw your attention to several things about it:

- We use the singular “is” and not the plural “are” to emphasize that statistics is a field of study, not just a bunch of numbers.
- We use “mathematical” as an adjective, because while statistics certainly makes use of much mathematics, it is a separate discipline and not a branch of mathematics.
- We use the noun “science,” for statistics is the science of gaining insight from data.

In this article we highlight some of the differences between statistics and mathematics, and we suggest some implications of these differences for teachers and students. We realize that these distinctions may not be universal but hope our broad strokes can help highlight some fundamental distinctions in the disciplines. Our aim is not to provide a philosophical discussion of these issues but rather to present ideas that will inform classroom practice. Toward this end we motivate and illustrate our points with concrete examples.

**Crucial Role of Context**

A primary difference between the two disciplines is that in statistics, *context is crucial.* Mathematics is an abstract field of study; it exists independently of context. Mathematicians often strive to “strip away” the context that can get in the way of studying the underlying structure. For example, one can study linear functions for their own mathematical properties, without considering their applications. Indeed, one could argue that worrying about the complications of real data diverts students’ attention away from the underlying mathematical ideas. But in statistics, one cannot ignore the context when analyzing data. Consider the following dotplot:
This plot reveals virtually nothing, it’s just a bunch of dots! Granted, we can see that one dot appears far to the left of the others, and there is a cluster of six dots less far to the left that seems to be separated from the majority of the dots. But we can’t interpret the plot or draw any conclusions from it. Let’s include a scale:

This is better, and we now know that the outlier is at 120, and the lower cluster is between 150 and 160. But we still can not gain any insights from this display. Let’s include the units:

Now we know that this graph reveals weights in pounds, but we still do not know whether we’re analyzing weights of statistics students or of horses or furniture or …

Finally, we’ll tell you the context: these are the weights of the rowers on the 2004 U.S. men’s Olympic rowing team. Armed with that knowledge, we can summarize what the dotplot reveals and even suggest some explanations for its apparent anomalies. It makes sense that the data include an outlier who weighs much less than the others: he’s the coxswain, the team member who calls out instructions to keep the rowers in synch but does not put an oar in the water himself. He needs to be light so as not to add much weight to the boat. The cluster in the 150’s also has an explanation: those six rowers participate in “lightweight” events with strict weight
limitations, two in a pairs event and the rest in a fours event. As for the rest, the majority of the rowers are big, strong athletes in the 180-230 pound range.

For another example of how context is paramount in statistics, consider the following scatterplot of data from a study about whether there is a relationship between the age (in months) at which a child first speaks and his/her score on a Gesell aptitude test taken later in childhood (found in Moore and McCabe, 1993). We have drawn the least squares regression line on the plot:

![Scatterplot](image)

This scatterplot and line reveal a negative association between the variables, indicating that a large value of one variable tends to appear with a small value of the other. Moreover, the slope of the line is statistically significant (p-value = .002, $R^2 = .410$). But on closer inspection, we see that this apparent negative association is driven largely by the two extreme cases in the bottom right of the plot. What do we make of this? Should we conclude that there’s a negative association here or not? Should we discount the outliers or not? Of course, to answer this question, we must first consider the context. We see that those two outliers are exceptional children who take a very long time to speak (3.5 years and a bit longer than 2 years) and who also have very low aptitude as measured by Gesell. To get a sense for whether the negative
association between speaking age and aptitude score holds for more “typical” children, we can delete the outliers and re-fit the line:

![Scatterplot with line](image)

This scatterplot and line reveal essentially no association between the variables. The slope coefficient is no longer statistically significant (p-value = .890, $R^2 = .001$).

So, our conclusion here contains two parts: Children who take an exceptionally long time to speak tend to have low aptitude, but otherwise there is virtually no relationship between when a child speaks and his/her aptitude score. (We should learn more about how these data were collected before deciding whether these results generalize to a larger group, and we could gather more data to examine whether those two exceptional children are indicative of a larger pattern.) Once again, the context drives our analysis and conclusions. Fitting a line to these data without considering the context would have blinded us to much of what the data reveal about the underlying question of speaking and intelligence.

**Issues of Measurement**

Another important issue that distinguishes statistics from mathematics is that *measurement issues* play a large role in statistics. Measurement is also important in mathematics; in fact, it is one of the standards in the NCTM Principles and Standards for School Mathematics (NCTM, 2000), but the focus is different. In mathematics, measurement includes getting students to learn
about appropriate units to measure attributes of an object such as length, area, and volume and to use formulas to measure those attributes. In statistics, drawing conclusions from data depends critically on taking valid measurements of the properties being studied. Measuring a rower’s weight is quite straight-forward, but measuring a child’s aptitude is quite challenging. Many other properties of interest in statistical studies of human beings are hard to measure accurately; examples include unemployment, intelligence, memory ability, and teaching effectiveness.

Another example of a hard-to-measure property comes from a fascinating recent study involving a city’s pace of life. Researchers studied whether a city’s “pace of life” is associated with its heart disease rate (Levene, 1990, as found in Ramsey and Schafer, 1997). They measured a city’s pace of life in three different ways:

- Average walking speed of pedestrians over a distance of 60 feet during business hours on a clear summer day along a main downtown street;
- Average time a sample of bank clerks take to make change for two $20 bills or to give $20 bills for change;
- Average ratio of total syllables to time of response when asking a sample of postal clerks to explain the difference between regular, certified, and insured mail.

The following scatterplots reveal that there is a slight positive association between heart rate and these three measures of a city’s pace of life:
Importance of Data Collection

One can study and do mathematics without analyzing data, but even when mathematicians examine data they typically focus on detecting and analyzing patterns in the data. How the data were collected is not relevant to purely mathematical analyses, but this is a crucial consideration in statistics. The design of the data collection strategy determines the scope of conclusions that can be drawn. Can you generalize a study’s results to a larger population? It depends on whether the sample was randomly selected. Can you draw a cause-and-effect conclusion from a study? It depends on whether it was a controlled, randomized experiment.

For example, consider two studies that asked women: “Do you give more emotional support to your husband or boyfriend than you receive in return?” In study A, 96% of a sample of 4500 women answered “yes,” but in study B 44% of a sample of 767 women answered “yes.” How do we reconcile these results? In which study do we place more confidence for representing the beliefs of all American women? The answers depend on how the data were collected. Study A was conducted by sociologist Shere Hite, who distributed over 100,000 questionnaires through women’s groups (Hite, 1987). Study B was conducted on a random sample of women, sponsored by ABC News and the Washington Post (Moore, 1992). Even with the smaller sample, study B provides more credible data because it involved a random sample.

For another example, consider two studies A and B that involve comparing “success” rates between two groups. The data are summarized in the table, including the p-value for comparing the two success rates with each study:

<table>
<thead>
<tr>
<th>Study</th>
<th>Group</th>
<th>Successes</th>
<th>Trials</th>
<th>Success rate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>42</td>
<td>61</td>
<td>.689</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>30</td>
<td>62</td>
<td>.484</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>806</td>
<td>908</td>
<td>.888</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>614</td>
<td>667</td>
<td>.920</td>
<td></td>
</tr>
</tbody>
</table>
The methods used for calculating these p-values are identical, and their values are very close. The small p-values indicate strong evidence of a statistically significant difference in success rates between the two groups. But does that mean that we draw identical conclusions from the two studies? The answer depends on how the data were collected.

Study A is a social experiment in which three- and four-year-old children from poverty-level families were randomly assigned to either receive pre-school instruction or not, with a response of whether they were arrested for a crime by the time they were 19 years old (as found in Ramsey and Schafer, 1997). Because the children were randomly assigned to a treatment, and because the p-value turned out to be so small, we can legitimately draw a cause-and-effect conclusion between the pre-school instruction and the higher success rate. On the other hand, study B is an observational study in which researchers examined court records of people who had been abused or not as children, comparing their rates of committing a violent crime as an adult (Widom, 1989, as found in Ramsey and Schafer, 1997). Because this was not a randomized experiment, no cause-and-effect conclusion between the child abuse and the violent crime rate can be drawn, despite the very small p-value. So while the calculations in two studies can be identical, the conclusions drawn can differ substantially, depending on how the data were collected.

**Lack of Definitive Conclusions**

Statistics and mathematics ask different types of questions and therefore reach different kinds of conclusions, with different thought processes to link questions to conclusions. Mathematics involves rigorous deductive reasoning, proving results that follow logically from axioms and definitions. The quality of a solution is determined by its correctness and succinctness, and there is often an irrefutable correct answer.
In contrast, statistics involves inductive reasoning and uncertain conclusions. Statisticians often come to different but reasonable conclusions when analyzing the same data. In fact, within these types of judgments lies the art of data analysis. All of statistical inference requires one to use inductive reasoning, as informed inferences are made from observed results to defensible, but ultimately uncertain, conclusions. In statistics we summarize conclusions with phrases such as “We have strong evidence that…” and “The data strongly suggest that…” but steadfastly resist saying things like “The data prove that…”. The quality of conclusions lies in the analysts’ ability to support and defend their arguments.

For example, we often ask students to collect data for comparing prices between two different grocery stores. This project raises many practical issues of measurement and data collection, such as whether students should record sale prices or regular prices, and how students should obtain a random sample of grocery items. Consider some sample data collected by our students of price differences between the two stores on 37 items:

How should we analyze these data? One reasonable approach is to calculate the mean of these differences and test whether it differs significantly from zero; a $t$-test yields a $p$-value of .308. Thus, the mean price difference in this sample does not differ significantly from zero. But we have not established that the average price is the same between the two stores; we have only concluded that the sample data do not provide compelling evidence to reject that hypothesis.

Not only is this conclusion uncertain, but we could have selected a different analysis altogether. We could instead perform a sign test of whether the median price differs significantly
from zero, which is equivalent to asking whether the proportion of items costing less in one particular store differs significantly from one-half. Twenty-one of the items are cheaper in one store, with only eight items cheaper in the other store (and eight “ties”). The p-value for this test turns out to be .024, which does suggest a statistically significant difference. But even this conclusion is uncertain, for the p-value reveals that sample data this extreme could have arisen by chance even if there was no difference between the median prices in these stores.

So, do the stores’ prices differ significantly, as the sign test suggests, or not, as the t-test suggests? Which of these two conflicting conclusions is correct? Which is reasonable? Well, neither and both. It depends on what question we want to ask (about the mean or median), and even then neither conclusion is certain. To complicate matters still further, statisticians might disagree on whether a normal model for the price differences is reasonable enough to justify applying the t-test, and statisticians might also disagree on how the “ties” should be handled when conducting the sign test. This lack of definitive conclusions, and even the lack of a single appropriate method of analysis, is common in statistical investigations.

**Communicating Statistical Knowledge**

Terminology is essential in mathematics as well as statistics, but one difference is that many common terms from everyday language have technical meanings in statistics. Examples include words such as bias, sample, statistic, accuracy, precision, confound, correlation, random, normal, confident, and significant. Students are very tempted to use these words loosely, without considering their technical meanings. Rather, studying statistics is akin to studying foreign language, for students need lots of practice to become comfortable using these terms correctly, and they often stumble at first before acquiring enough familiarity to use the language well. Although there are English terms such as “multiple” and “factor” with a technical meaning in
mathematics, it seems that students expect to see more technical terms in mathematics than in statistics.

Communication is essential in both disciplines, but more so in statistics because it is a consulting enterprise. Statisticians routinely must interact with clients whose technical skills vary greatly, from eliciting a clear statement of the problem from those clients through communicating to them the results and conclusions of the analysis. While introductory students are far from professional statisticians, the ability to communicate statistical ideas in layperson’s terms is essential and an important component of many courses. Communication is important in mathematics also, but that communication is more often done symbolically in mathematics.

**Why Should Teachers Care?**

We have argued that statistics is a different discipline than mathematics, that it involves a different type of reasoning and different intellectual skills. Even if you find our case persuasive, the question remains: Why should classroom teachers care? We see two primary reasons:

- A different type of instructional preparation is needed for teaching statistics.
- Students react differently to statistics than to mathematics.

In order to help students see the relevance of context, measurement issues, and data collection strategies in statistics, it’s imperative that teachers present real data, in meaningful contexts, from genuine studies. Fortunately, there are a plethora of resources available now to help teachers with this, from books to CD-ROMs to websites (Moore, 2000).

Instructors also need to help students learn to relate their comments to the context and to always consider data collection issues when stating their conclusions. While this is sometimes done in mathematics courses, it’s not nearly as prevalent or as essential as it is with statistics. Indeed, this need for different types of assessments is a key difference between teaching statistics
and mathematics. Many students do not initially expect this type of focus and teachers need to be prepared for students’ discomfort. Students also need to be reminded that there are multiple correct approaches, and that they will also be evaluated on how well they explain and support their conclusions.

Another difference in terms of instructional preparation is that many teachers do not have ample opportunities to develop their own statistical skills and understanding of statistical concepts before teaching them to students. This challenge is especially acute because few programs in mathematic teacher preparation offer much instruction in statistics, and much of the instruction that is provided concentrates on the mathematical aspects of statistics. The recent Mathematical Education of Teachers report makes these points quite forcefully (Conference Board of the Mathematical Sciences, 2001). Helping students to develop their communication skills and statistical judgment, so crucial in the practice of statistics, is also very challenging and an area in which mathematics teachers are provided with little instruction.

The experiences and reactions of students to studying statistics are different from studying mathematics. Educational research shows that students (and others) have tremendous difficulties with reasoning under uncertainty (Garfield, 1995; Shaughnessy, 1992; Garfield & Ahlgren, 1988). Also, many students (and others) are very uncomfortable with uncertainty, with the lack of definitive conclusions, and with the need for detailed interpretations and explanations that are integral to studying statistics. Helping students to develop a healthy skepticism about numerical arguments, without allowing them to slip to the extremes of cynicism or naïve acceptance, is a great challenge.

Because of the differences between statistics and mathematics, teachers should expect that some mathematically strong students may be frustrated while studying statistics. But on the
bright side, many students who may not be initially excited by mathematics will be intrigued and empowered by their experiences with statistics.

References


