Michigan Council of Teachers of Mathematics

Mathematics in Michigan

Volume 46, No. 2

Fall 2013

Common Core State Standards for Mathematical Practice

Reason Abstractly and Quantitatively

A scholarly journal for mathematics educators in Michigan

Mathematics in Michigan MIM

Volume 46, Number 2,
Fall 2013

Features

Departments

Using Data from Real Sources

Christopher Shaw and

Gregory Johnson

Standard 2 of the *Standards of Mathematical Practice* calls for students to develop abstract and quantitative reasoning. The goal of this project is to use a concrete in-class example to lay the foundation for a thorough understanding of several basic statistics concepts, in order that students may gain confidence when applying the definitions in an abstract setting.

INTRODUCTION

On a typical first day of an elementary statistics course, one spends much of the time hoping to motivate the main concepts of *center* and *variation* for a given data set. Many students will already be familiar with how to calculate the *mean*, *median*, and *mode*, of a finite list of numbers, but may not have an intuitive understanding of why these different measurements of center exist, nor how, given an arbitrary data set, they can be interpreted. In order to properly motivate the definitions, the first several chapters or sections of an elementary statistics textbook usually begin by establishing a *hypothetical* data set with behavior that helps to clarify the meaning of each of the statistical measures. However, our experience indicates that students in an elementary course will be more likely to resonate with data that are collected from *real sources*. Fortunately, a classroom full of students is a perfect environment in which to dismiss the hypothetical and create from the class an actual example with exactly the desired properties. A quick survey can accomplish this goal, with the added benefit that it requires direct participation in order to be successful.

This project is intended for implementation in grade levels 13-14. The activities were initially worked out in a college-level elementary statistics course, and in a terminal college-level introductory mathematics course typically used to meet the basic math requirement of a liberal arts curriculum.

THE MOTIVATING SURVEY

Preparation

For this survey, the students in the class were each handed a slip of paper and asked to answer the following four questions:

- Guess your teacher's age, in years.
- What is your height in feet, to the nearest thousandth?
- How many tattoos do you have?
- What is your class in school (*i.e.* Freshman, Sophomore, Junior, Senior)?

The rationale behind each of these questions is discussed below, but instructors are encouraged to personalize or other questions that have the desired properties. The results listed in Table 1 represent the responses for a typical class.

Data collection

Our suggestion is to pass all of the survey results to a student, and ask the student to read the responses aloud as the instructor tabulates them on the board. The compilation of survey data carries a lesson in and of itself, as the most efficient method will most likely be a frequency table.

LESSONS

Center

To encourage a discussion about what the students already know, the instructor may ask the class to speculate about the meaning of the word *average*, and how they would find the right value to use for the colloquial average for each of the properties measured. (In our experience, many students will instinctively calculate the arithmetic mean of the numerical responses for each of the questions, and assume this is the intended answer. After getting input from the students, the instructor can spend a few mo-

1. Teacher's Age (years)	2. Height (feet)	3. Tattoos	4. Class
26	5.417	0	Freshman
28	5.083	1	Freshman
29	5.083	0	Sophomore
30	6.041	0	Freshman
31	5.500	0	Freshman
37	5.083	0	Junior
36	5.833	2	Senior
28	5.750	1	Freshman
27	5.500	0	Sophomore
36	5.917	1	Sophomore
35	5.958	1	Freshman
30	5.875	0	Freshman
31	5.250	1	Sophomore
32	5.833	4	Senior
30	5.167	0	Freshman
31	6.000	0	Freshman
36	5.917	0	Freshman
25	5.583	0	Sophomore
24	5.333	0	Freshman
29	5.333	1	Freshman
30	5.750	2	Sophomore
31	5.167	1	Junior
32	5.875	0	Freshman

Table 1

ments going over calculations for the mean, median, and mode. This is also a good time to point out that the frequency table makes finding the modes trivial, and makes finding the median nearly so: to find the mode(s), one need only find the value(s) of highest frequency, and to find the median, one ensures that the data values are listed in order, and counts along the frequencies from bottom to top to select the middle value from the list. As we will describe below, the advantage to using the questions above is in the fact that each of them has properties that make it a good candidate for a particular choice of average.

Guess my age: In our experience, the answers to this question are approximately normally distributed, with mode values near the center of the data. In the above example the *mean* is about 30.6, the *median* is 30, and the *modes* are 30 and 31. Based on their proximity, one can argue that any of the three candidates is a good choice of representative for the data. It is pertinent to note here that in the example data reproduced here, the instructor's actual age was close to the mean of the guesses, which may not be the case if this survey is used for an instructor whose age is further from that of the students in the classroom. One may expect a wider variation in guesses as the age gap between the students and the instructor increases.

Height: The authors have used this data set in the classroom to illustrate one main point. In a small data set with a large possible range in response values, the mode is often a meaningless measurement. In these responses, the *mode* is 5.083 feet, a value shared by three of the respondents, but it does not represent a central value in the data (in fact it is the minimum response). As such the mode seems a poor choice for *average* in this category.

Tattoos: The answers to this survey question are likely to be different depending on the demographics of the audience involved; the second author most recently used this survey for students attending an arts and media college, and without making too strong a generalization, the data from these tiny samples indicate that those students were much more likely to have tattoos than their counterparts at a large state school. However, even in this case the results were useful in demonstrating that the mean is not always a good measurement of *average*. In particular, with a mean of about 0.65, neither the mean nor the mean rounded to the nearest whole number would be sensible averages for the class, as over half of the students had no tattoo at all. This fact is reflected by both the median and the mode being zero.

	1. My age	2. Height	3. Tattoos	4. Class
Standard Deviation	3.5	0.32	0.96	0.95
Relative Standard Deviation	11.4%	5.8%	147.3%	136.9%

Table 2

Class: The instructor should emphasize here that the data collected in the final question of this survey is not numerical, and thus there is no well-defined way to calculate a measure of center. However, it is clear that there is still a quantitative bent to the question, in that the possible responses have an inherent ordering. In this way the question differs from a strictly categorical one asking, for instance, for a student's favorite food. The authors note that in general, the idea of a student's year in school is not a rigorous notion: the definitions may vary from school to school, and depend on the traditional four-year model in order to make sense. However, if one were interested in gauging the average amount of time that students in the class had spent in college, the data collected could be used to make a very rough estimate, with some adjustment in the form of conversion to numerical values. An effective conversion is to map the classes to the numbers 0 – 3, in which case the mean is 0.7. This way, the representation defines the number of full school years a student in each category has completed, allowing for a sensible summary sentence: "On average, students in this class have completed seven tenths of a year of college."

Variation

In our discussion of standard deviation, we find it helpful to put off the description of the mathematical formulae used in calculation for as long as possible. Instead, we begin by introducing standard deviation as "a measure of the average variation from the mean of the measured responses." This concept, while not a rigorous definition, has the advantage of being both correct and slightly ambiguous, which allows for a discussion later of how one might go about calculating such a value. Of course there are many measurements one might take that meet these criteria, and often it is not until learning more advanced geometry (namely, a rationale for the distance formula) that a student will understand why the actual standard deviation is, in fact, ideal.

To begin, the authors ask student participants to look at the results and ask which of the columns appears to have the most uniform set of responses. In our experience, the tattoo column, with its overwhelming mode of zero, will draw the students to suggest that column as the most uniform. And in fact, once standard deviation is introduced, this column will almost certainly have a low standard deviation value. But it is easy to use these datasets to convince the students that simple variation is not by itself an appropriate measure of uniformity. Ask the question, "Which is more significant, the difference between a 30-year-old and a 31-year-old, or the difference between having two tattoos and having three tattoos?" This question will help to solidify the idea that the context matters. This allows the instructor to lead the class into a discussion of both the standard deviation, which measures the raw average variation in data, and the relative standard deviation, which divides the standard deviation by the mean to give a depiction of uniformity that can be compared across datasets.

Depending on the structure and pace of the class, the instructor may or may not be interested in asking the students to calculate either of these statistics. Regardless, the class will have heard of the term *standard deviation*; as such, introducing it as a black box that satisfies the definition above should not cause confusion. Standard deviations and relative standard deviations for each of the data columns are included in the table above.

In comparing these data, for instance, the standard deviation for the age data is much larger than that of the tattoo data. Yet, when we look at the responses to the age question, they seem to be distributed nicely in a cluster around the center. In contrast, the standard deviation for the number of tattoos seems small, at nearly 1. However, when this value is divided by the mean response, the relative value is much higher, making it clear that the tattoo question has the most relative variation among responses.

FURTHER USE

The authors continue to refine the survey in order to balance the pedagogical value with the amount of class time used on introductory topics prior to learning techniques of calculation. As this process continues, it has become evident that this simple survey can be reused throughout the course. While the scope of this article is limited to use in the first day of a new class, as the semester progresses into more advanced topics, one may consult the survey data in order to help explain, among other topics, left and right-skewed distributions (as in the tattoos), bimodal distributions (as in the height), and the concept of continuous vs. discrete variables (by allowing arbitrary decimals for the height).

Christopher Shaw is Assistant Professor of Mathematics at Columbia College Chicago. His mathematics research is in logic, and his pedagogical focus is on entry-level college mathematics. cshaw@colum.edu

Gregory Johnson is Visiting Assistant Professor of Mathematics of Carnegie Melon University. His mathematical research is in model theory. greggo@math.cmu.edu