Means and MADs

GARY D. KADER

IN STATISTICS, WE COMMONLY EXAMINE variation in data from the perspective of how the data values differ from a representative value. If the mean is used as a representative value, we ask, "How different are the data values from the mean?" The following activity builds a framework for developing students' understanding of the notion of variation from the mean. The activity is a modification and an extension of one described by Friel, Mokros, and Russell (1992) and is designed to develop the following statistical concepts:

- The distribution for a set of data
- Various interpretations of the mean for a set of data
- The limitation of the mean for distinguishing distributions
- Variation in data
- Variation in data from the mean
- The MAD as a measure of variation in data from the mean

The Problem

EIGHTH-GRADE PREALGEBRA STUDENTS WERE given the following problem:

> Nine people were asked, "How many people are in your family?" One result from the poll is that the average family size for the nine people was five.

First, I showed students one possible result for the nine family sizes **(fig. 1)** and asked them to interpret and comment on these results. Most students agreed that this distribution indicated that each person in my survey had five people in her or his family and that if all nine people polled had families of size 5, the mean would be 5. They further concluded that although all nine people could have families of size 5, it was unlikely that this distribution would happen very often. Certainly, many distributions of family sizes have a mean of 5, which leads to another question:

> What are some other possibilities for the distribution for the nine different family sizes with the requirement that the mean family size be 5?

Working in groups, students used sticky notes to form line plots displaying the nine people's family sizes with a mean family size of 5 **(fig. 2).** To simplify the problem, the smallest family size allowed was 2 and the largest family size allowed was 11. Each group was given nine sticky notes and a sheet of poster board with an axis scaled from 2 to 11. The groups were encouraged to be creative in forming their distributions. For example, I allowed only one or two groups to have symmetric distributions. In each group, students were required to write a description of their strategy. Some of the students' distributions are shown in **figure 3.**

Many students relied on their knowledge of the algorithm for finding the mean to form their distributions. Often they began by examining their own family sizes or by arbitrarily placing sticky notes over several different values and then putting the remaining sticky notes over the values needed to have the total of the family sizes be 45. For example, two groups described their strategies as follows:

> What our group did was guess and check. We just took a big group of nine numbers and added them up. We kept changing the numbers around until we got 45.

> In my group our strategy was to figure out how many times 9 will divide into a number to equal 5, then we picked, randomly, nine numbers. We added those together and divided by 9. Our answer was high, so we lowered some of our numbers until they all equaled 45. Then we knew that we had the right answer.

GARY KADER, *gdk@math.appstate.edu*, *teaches at Applachian State University, Boone, NC 28608. He is involved in teacher education and the development of learning activities and has a particular interest in statistics education.* *The author thanks Karen Hensley and her eighth graders at Blowing Rock School, Blowing Rock, NC 28605.*

```
            X
            X
            X
            X
            X
            X
            X
            X
            X
--+---+---+---+---+---+---+---+---+---+--
  2   3   4   5   6   7   8   9   10  11
       Mean Family Size Equals 5
```

Fig. 1 One possible result for nine families

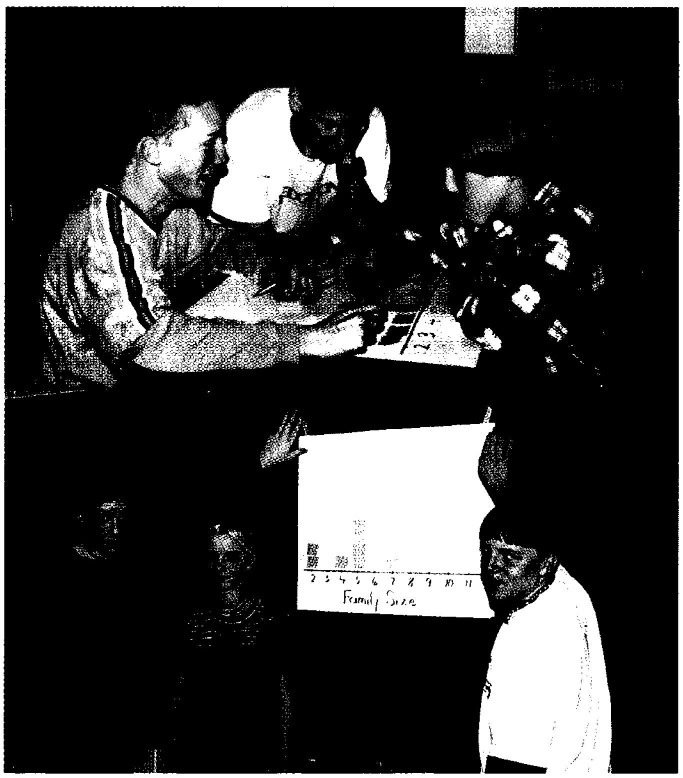

Fig. 2 Students formed line plots by using sticky notes.

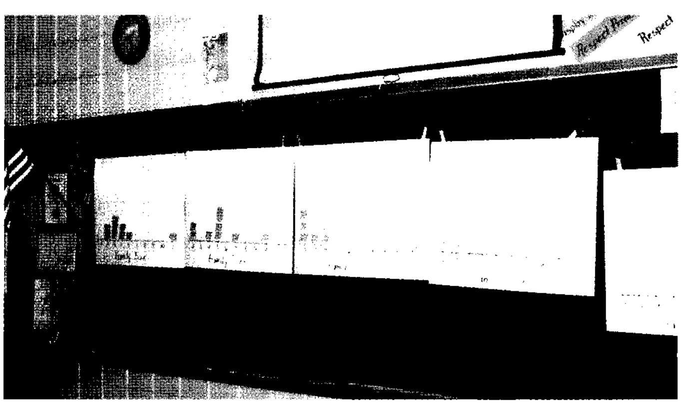

Fig. 3 Line plots showing students' distributions

Variation from the Mean

THE OBJECTIVE OF HAVING STUDENTS GENERATE DIStributions was to allow them to see that different distributions can have the same mean. Because their distributions were not adequate for developing the desired objectives of this activity—measuring variation from the mean—students were given distributions that showed greater variability. The eight distributions shown in **figure 4** were presented to the class with the same type of sticky-note-and-poster-board display. Each distribution has a mean of 5, but the distributions vary greatly. Students were asked to discuss the limitations of knowing only the mean family size. Some student comments follow:

> If you only have the mean, it limits you to not knowing how many families are at each of the numbers or how many people are in each family.
>
> It limits what I know by not telling you what the other family numbers are.

Thus, knowing only the mean does not provide enough information to identify which distributions actually occurred.

I posed the following question to the class for discussion:

> What additional information about the data could be given to identify which of the distributions in **figure 4** matched the results from my survey?

The line plots in **figure 4** illustrate the pattern of variability for each distribution. A major goal of statistics is to offer ways to summarize and measure this variability. One approach is to consider by how much the individual measurements differ from the mean. Students were asked to respond to the following two questions:

> Of all these distributions, which shows data values that differ the least from the mean value, 5? Explain.
>
> Of all these distributions, which shows data values that differ the most from the mean value, 5? Explain.

All students recognized that distribution 1 has data values that differ the least from the mean of 5. Some of the better explanations follow:

> All of the families were made up of five members, which was the mean.
>
> Graph 1 because all of them are on the same number.
>
> The one with all 5s differs the least because it is showing you flat out that a family has five family members.

Most students recognized that distribution 8 has data values that differ the most from the mean of 5. The following explanations show thoughtful analysis:

This is because the only families were made of 2 and 11. None of these were close to 5.

The chart with six families on 2 and three families on 11 because 2 and 11 are the farthest from 5.

This is so because no families are on 5 or anywhere near 5.

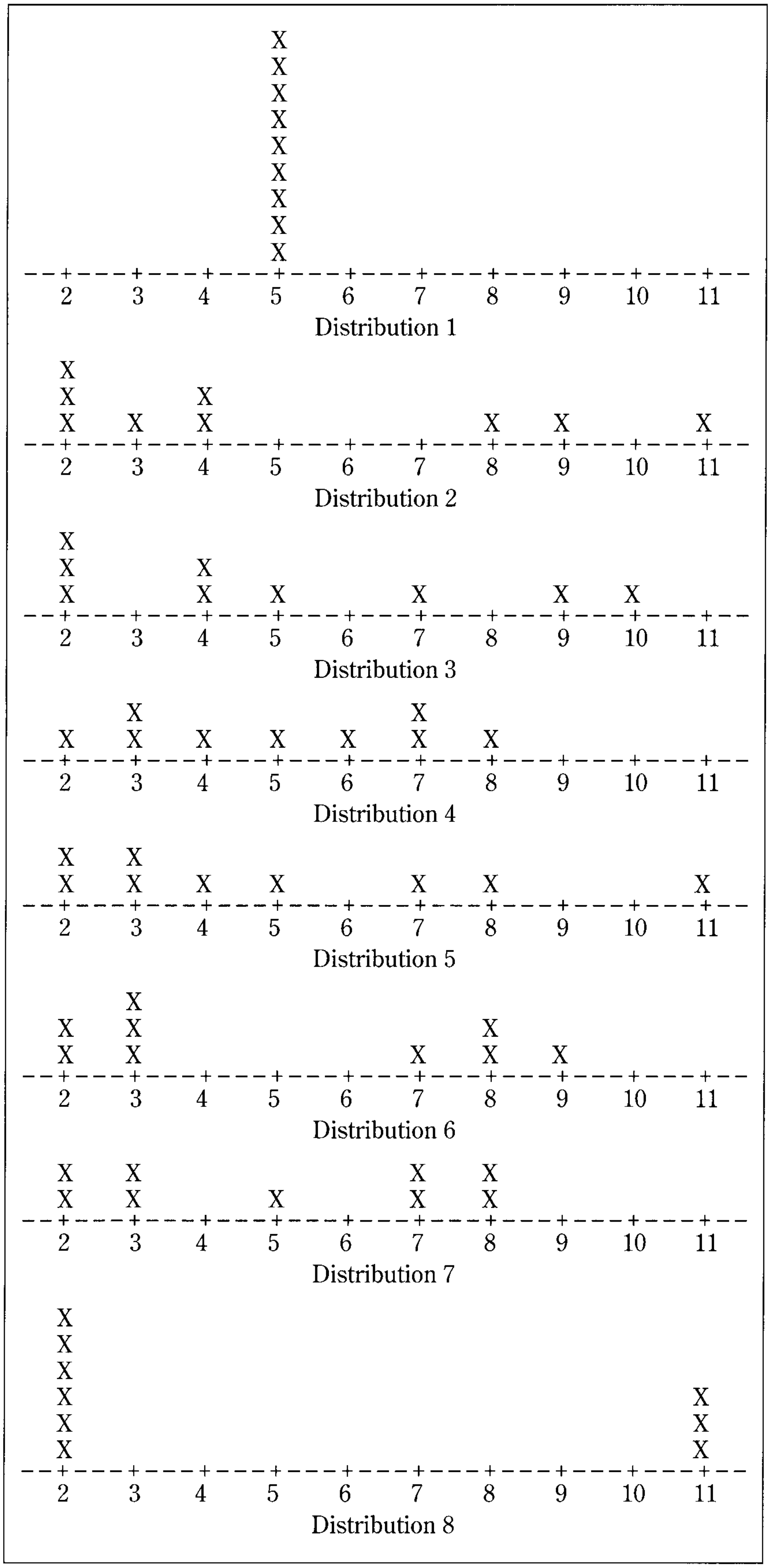

Fig. 4 Eight distributions: Mean family size equals 5 for each distribution.

The students were then asked to consider the following problem:

> On the basis of how different the data values appear to be from the mean of 5, how would you order, from least to most, the other six distributions (distributions 2–7)? Make a list of the order in which you think these distributions should be. Explain your reasoning.

Each student determined an ordered arrangement of the distributions in **figure 4** on the basis of a visual examination. All students had difficulty expressing their reasoning, and none adequately explained how he or she had arrived at the ordered list. The class used the individual orderings to vote on a consensus ordering for the class. The class decided on this order: 4, 7, 3, 5, 2, 6.

Although some students' lists matched the one selected, many disagreed with the final list. The difficulty with ordering the distributions through a visual examination is its subjectivity. Instead of forming a subjective ordering, students need to quantify the magnitude of these differences. Quantification offers an objective mechanism for making a decision about how different the data values are from the mean. One quantity that the class examined was the difference between each data value and the mean. This difference is called the deviation of a value from its mean, or simply the deviation from the mean:

Deviation from the mean = value – mean

Students determined the deviations for each distribution in **figure 4** and displayed them on the line plot by writing the deviation on the sticky note over each data value. The deviations for distribution 2 are displayed in **figure 5.** One observation from distribution 2 is that the deviations add to 0. Will this result always be true? Students added the deviations for each distribution in **figure 4.** In each instance, these deviations sum to 0. In general,