Students' Reasoning about Summary Measures from Histograms and Stem-and-Leaf Plots

Linda Cooper, **Felice Shore**
Mathematics Department, Towson University, 8000 York Road, Towson, Maryland, 21252

Abstract

This paper identifies and discusses misconceptions that students have in making judgments of center and variability when data is presented graphically. An assessment addressing interpreting center and variability in histograms and stem-and-leaf plots was administered to undergraduates enrolled in upper and lower level introductory statistics courses. In particular, discussions focus upon comparing the variability of two sets of data having common mean, median, and range, represented by bell-shaped histograms of common scale, and in comparing the relative position of the mean and median on a positively skewed histogram. When the exact values of the data were available in stem-and-leaf plot representation, students were more successful at computing the mean and median values.

KEY WORDS: misconceptions, histograms, center and variability

1. Introduction

Introductory college statistics courses typically spend little time on descriptive statistics, presumably because the perceived goal of these courses is to provide students with a basic understanding of inferential statistics. The mean and standard deviation quickly become central to more complex concepts and formulas. As students move beyond descriptive statistics, our concern is that little time has been spent making connections between measures of center and variability, and graphical representations. An implication, for example, is that the conceptual groundwork for student understanding of the Central Limit Theorem may be compromised by the inappropriate presumption that students can extract meaning from a histogram with regard to the mean and variation.

Based upon experiences of pilot studies, it is our contention that students actually have a very tenuous understanding of the mean and median beyond computation, that students have much more trouble interpreting basic graphical representations than might be presumed, and that in general, they have few notions with which to reason about quantitative data when provided in the aggregate, as in histograms. In the present study, we attempted to catalogue students' ways of reasoning about and computing measures of center and variability from histograms and stem-and-leaf plots.

2. Methods

2.1 Sample

For this study, researchers administered a 15-minute assessment on statistical representations and concepts to 186 students in nine sections of statistics courses at a large public university. Forty students were enrolled in 300-level statistics courses in the Mathematics Department. The remaining 146 students were distributed among lower-level statistics courses in the Mathematics and Psychology Departments. The assessment was administered only after all descriptive statistics course material was completed.

2.2 Instrument

The authors developed a four-item assessment based upon previous pilot research. Items were chosen for their indicated potential to reveal student thinking about topics of interest. Items 1 and 2 are multiple choice and require students to correctly interpret histograms of grouped data. Both items 3 and 4 ask students to compute the mean and median from graphs with accessible raw data.

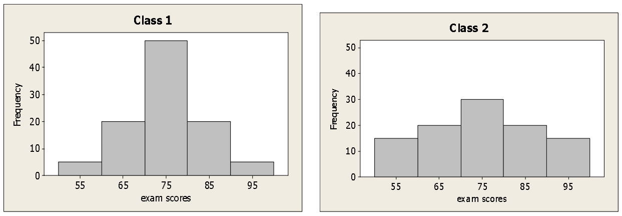
2.3 Analysis

Once all data were collected, multiple-choice responses to items 1 and 2 were entered into a database, as were numerical responses to items 3 and 4. In addition to recording responses, for items 3 and 4, the two researchers also coded student methods of reasoning based on work shown. In a future research effort the authors intend to conduct student interviews to a) corroborate our interpretations of student thinking based on their written responses and b) further understand how students think about the graphical representations in ways that are not possible to deduce from the written record.

3. Results

The first item (Figure 1) assessed students' ability to compare the variability of two sets of data sharing the same mean, median, range, and bell-shape distribution, represented by histograms of common scale.

1. The following graphs show the distribution of exam scores in two classes.



Comparing the two distributions, one could infer

i. class 1 had greater variability than class 2.
ii. class 2 had greater variability than class 1.
iii. class 1 and class 2 had equal variability.
iv. I don't know.

Figure 1: Assessment item 1

As indicated in Table 1, 27.4% correctly responded that exam scores of class 2 had greater variability than exam scores of class 1. Roughly half of the students responded that exam scores of class 1 had greater variability than those of class 2. One interpretation of this response is that variability of the data is being judged by the variability of the heights of the bars. In that case, the histogram with narrow tails and a high peak indicate great variability; whereas, the histogram with bars of more similar heights indicates little variability.

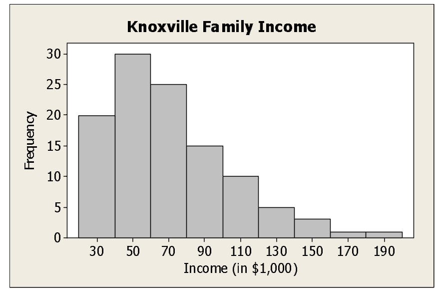
Table 1: Distribution of Responses for Item 1

Response	n	(%)
i.. var_{Class1}>var_{Class2}	92	(49.4)
ii. var_{Class2}>var_{Class1}	**51**	**(27.4)**
iii. var_{Class1}=var_{Class2}	37	(19.9)
iv. I don't know	5	(2.7)
other	1	(0.5)
Total	186	(100.0)

The 20% who judged the variability to be equal may simply be judging variability by the range of the data set. Later items indicated that students tend to overly focus on the horizontal scale. Since the range of the data is easily gleaned from the horizontal axis, it is conveniently extracted as the measure of variability, crude and unsophisticated as it may be, for comparison purposes.

On item 2 (Figure 2), students were given a positively skewed histogram and asked to compare the relative positions of the mean and median. Table 2 provides the results for that item.

2. A study was conducted to examine the standard of living for typical families in Knoxville. The following graph displays the distribution of family income for those in the town of Knoxville.



Which of the following statements is a correct comparison of the mean and median family income in Knoxville?

i. The mean income is less than the median income.
ii. The mean income is equal to the median income.
iii. The mean income is greater than the median income.
iv. It is impossible to determine which measure is larger from the given graph.
v. I don't know.

Figure 2: Assessment item 2

Thirty-two percent of the students correctly responded that the mean was greater than the median. Twenty-eight percent stated that the mean was less than the median, while 14% stated that the two measures of center were equal.

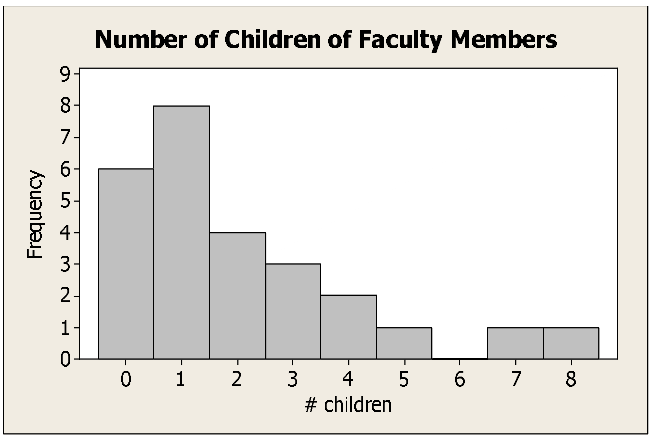
Table 2: Distribution of Responses for Item 2

Response	n	(%)
i. mean < median	53	(28.5)
ii. mean = median	26	(14.0)
iii. mean > median	**60**	**(32.3)**
iv. Can't tell from graph	29	(15.6)
v. I don't know	14	(7.5)
no response	4	(2.1)
Total	186	(100.0)

Some students may have experience comparing the relative measures of center for skewed and symmetric distributions, or be able to reason what effect a tail has on each of these measures of center. These students should have been able to answer this question without having to perform any calculations whether they understood the reason why or merely memorized the fact that a mean is on the tail side of the median in a skewed distribution. Without experience or knowledge about the effect of the tail, one may choose to approximate the value of the measures of center in order to make the comparison. However, the researchers observed virtually no written work on the assessment papers and very little calculator use for this problem. This could indicate that students may have visually estimated the locations of the mean and median in order to compare the measures. If that is the case, student work on subsequent item 3 offers a clue as to why they might have incorrectly positioned the mean and median. After detailing results from item 3 below, we will return to this discussion.

Like multiple choice item 2, item 3 (Figure 3) also presented a positively-skewed histogram, except in this case the data were ungrouped, leaving the raw data completely accessible. For item 3 students were asked to actually compute the mean and median, and their subsequent written work provided a means to categorize common errors.

3. The following histogram shows the number of children of faculty members in the Mathematics Department.



a. Find the median number of children of faculty members.

b. Find the mean number of children of faculty members.

Figure 3: Assessment item 3

As indicated in Table 3, 46% of the students were able to correctly determine the median value to be 1. Forty-four percent correctly found the mean to be approximately 2.04.

We examined the written supporting material that many students included to support or explain their reasoning. Patterns of errors emerged. Most notably was the failure to maintain the link between the values on the horizontal axis and their corresponding bar height or frequency.

Table 3: Distribution of Responses for Item 3a

Response	Interpretation	n	(%)
1	**correct median**	**85**	**(45.7)**
2 or 2.5	median height of bars	13	(7.0)
3.5, 4, or 4.5	median or midrange of horizontal axis	66	(35.5)
other		22	(11.8)
Total		186	(100.0)

On item 3a, 36% of the students (n=66) found the median value to be either 3.5, 4, or 4.5. Forty-six of these sixty-six students provided no written work to lend insight to their reasoning. Of the 20 students who showed work on the written assessment, fifteen demonstrated that they found the median by finding the midpoint of the values listed on the horizontal axis. There were several variations of this theme. Thirteen students either a) listed the values 0 through 8, b) listed the values 0 through 8, omitting the value 6 because its frequency was 0, or c) used the existing values on the axis and then crossed off values to find the middle value, arriving at an answer of 4 or 3.5, depending on whether they excluded the value 6. Two students mistakenly indicated that since there were nine values, the median would occur at 4.5. These students confused the position of the median (4.5) with the value of the median (3.5) from the data set 0, 1, 2, …, 8. Two other students calculated the midrange of 0 to 8 to be 4 by showing the expression for summing 0 and 8 and dividing by 2. Finally, three students constructed a frequency table, with no additional written work that could justify their response.

Approximately 7.0 percent (n=13) of the full sample found the median to be 2 or 2.5. Some of those students included work that helped us understand their solution method. Eight of these students listed the values of the height of the bars in order, and found the median of this list of values. One student confused the median with the mean, and incorrectly reported the average of 2 to be the median.

Parallel misconceptions were found on item 3b (Table 4). Whereas the most common misconceptions on item 3a were to find the median value of the horizontal axis, or median value of the heights of the bars, the most common misconceptions on item 3b were to find the mean value of the horizontal axis and mean value of the frequencies. Nineteen percent (n=35) mistakenly found the mean of the distribution by finding the mean value of the values on the horizontal axis. Thirteen percent (n=25) found the mean of the distribution by finding the mean value of the heights of the bar. Responses were grouped by method, disregarding slight variations. For example, responses for the mean value of the horizontal axis varied depending upon inclusion of "6" whose frequency was 0, and division by either 8 or 9 [36/9=4, 36/8=4.5, 30/8=3.75, 30/9=3.3]. Responses for the mean value of the bar heights varied depending upon whether the student included the value of 6 as a potential bar [26/9=2.89, 26/8=3.25].

Table 4: Distribution of Responses for Item 3b

Response	Interpretation	n	(%)
2.00 - 2.04	**correct mean/ correct method with arithmetic errors**	**72/81**	**(38.7/43.5)**
2.80 - 3.25	mean height of bars	25	(13.4)
3.30 - 4.50	mean of horizontal axis	35	(18.8)
other		45	(24.3)
Total		186	(100.0)

In summary, consistent patterns of errors emerged as students calculated the mean and median from the histogram of ungrouped data. The most frequent responses for both 3a and 3b were the correct response [45.7% and 43.5% respectively]. The most frequent misconception in finding the median / mean of data represented via a histogram was to find the median / mean of the values listed on the horizontal axis without regard to the height of the bars above [35.5% / 18.8%] followed by finding the median / mean of the frequencies of the data values [7.0% / 13.4%]. The former misconception might help explain students' incorrect estimates that would lead to placing the mean less than the median in item 2. If students used only the values on the horizontal axis to identify the middle (median) of the data, they would have a "median" value [110] much greater than it actually should be. These same students might then have compared the "weightiness" of values less than the "median" and the small tail greater than the "median", and concluded that the mean of this positively skewed graph would be less than the median.

Finally, we conjectured that a greater percentage of students would be able to find the mean and median of a data set represented by a stem-and-leaf plot than for a similar data set represented by a histogram. Students familiar with a stem-and-leaf plot are aware that the raw data values are easily retrievable. Though this was also the case with the ungrouped data in the histogram of item 3, it is not the case with histograms using grouped data. On item 4 (Figure 4), students were presented with a positively skewed stem-and-leaf plot showing the ages of patrons in a restaurant at a particular time.

4. The following stem-and-leaf plot displays the ages of patrons in a restaurant at a particular time.

1	0 2 8
2	1 1 3 7 8
3	0 4 6
4	5 5
5	2 8
6	
7	8

Key
2 | 3 indicates 23 years

a. Find the median age of patrons at the restaurant.
b. Find the mean age of patrons at the restaurant.

Figure 4: Assessment item 4

Fifty-two percent successfully found the median (Table 5) and 62% (Table 6) successfully found the mean. Far fewer students showed their work for item 4 and in fact more students left these items blank than any other, possibly a sign of assessment fatigue. Many of the responses were close in value to the correct answers and quite likely were found using the correct method with arithmetic errors. As it was impossible to tell if the "close values" indicated a correct method with arithmetic error or an incorrect method, in the case of the mean, only responses with the exact answer of 33.625, or rounded to either 33 or 34 were accepted as correct. This is a deviation of scoring from item 3 where student work allowed the authors to more easily categorize the responses. Responses from the written assessment did not indicate a clear pattern of misconceptions, although there were more explicable errors in finding the median than the mean. One noted error was ignoring the meaning of the stems and reporting the middle leaf to be the median. Another error was identifying the middle stem and either reporting that number (4) or using the age for that stem (45).

Table 5: Distribution of Responses for Items 4a

Response	n	(%)
correct median: 29	**96**	**(51.6)**
no response	18	(9.7)
other	72	(38.7)
Total	186	(100.0)

Table 6: Distribution of Responses for Items 4b

Response	n	(%)
correct mean: 33 to 34	**115**	(61.8)
no response	23	(12.3)
other	48	(25.9)
Total	186	(100.0)

A comparison was made across related items (Tables 7 and 8). The assessment presented items in order of how accessible raw data was from the graphical representation. Item 2, which presented grouped data in a histogram, made raw data impossible to retrieve; items 3 and 4 involved graphical representations of raw data values, although those values were more readily apparent in item 4. It is also arguable that the graphs were presented in order of most to least sophisticated. Whether due to the decrease in complexity of the items, or increasing familiarity with the topic, students performed better with each subsequent item for a given topic.

Table 7: Percentage of Correct Responses for Problems Involving the Median

Item	correct responses n (%)
2a: comparison of mean and median [histogram with grouped data]	53 (28.5)
3a: calculating median [histogram with ungrouped data]	85 (45.7)
4a: calculating median [stem-and-leaf plot]	96 (51.6)

Table 8: Percentage of Correct Responses for Problems Involving the Mean

Item	correct responses n (%)
2a: comparing mean and median [histogram with grouped data]	53 (28.5)
3b: calculating mean [histogram with ungrouped data]	81 (43.5)
4b: calculating mean [stem-and-leaf plot]	115 (61.9)

4. Conclusions

Our survey has revealed several insights about the link between student understanding and student performance on questions about statistical graphs and concepts. One finding is that students may be able to answer some basic questions about histograms without fully understanding how the distribution of the data links the frequencies (heights of bars) with values on the horizontal axis. When confronted with the question of computing measures of center from this type of graph, difficulties arose, particularly with interpreting numbers resulting from intermediate calculations, keeping the context of the numbers in mind. Whereas most students had a strong connection between median as "middle," it was clear that many misunderstood what middle value they needed to find when the data were summarized in a graph. That is, they either lost track of or were unaware of which numbers represented data values (in contrast to which numbers represented frequencies or scale values).

The item asking students to compare variability in histograms indicated that students' notions of variability are indeed tenuous. First, students are initially and appropriately taught that range is a measure of variability. This crude measure is easily gleaned from a graph – much more so than standard deviation. Thus, it is not so surprising that 20% of students would use it as the sole measure to assess comparison in variability. The more troubling finding is that students judged variability by focusing on the varying heights of the bars, implying variability in frequencies, rather than data values. A possible source of this confusion may be the surface-level visual similarities between ubiquitous bar charts, time-plots that use bars, and histograms, as the methods to evaluate variability differ dramatically for these different types of graphical representations.

5. Implications

Our research suggests the following implications for instruction:

Instructors should explicitly discuss the concept of variability of data in general and not limit the focus to quantifying variability through common measures such as range, interquartile range, and standard deviation. We want students to have a sense of what is meant by variability of data. It is important to acknowledge that the concept of variability is inherently more abstract than that of center. Whereas one can estimate a measure of center, it is not so easy to approximate variability other than range. More time needs to be spent developing the concept of variability within the context of data presented in different kinds of graphs.

To gain a better understanding of how variability is represented in histograms of quantitative data, students should examine histograms of little and great variation. One possibility is to have students start with a "discrete uniform" distribution where all bars have the same height. A discussion would follow that focuses on how distributions with the same mean and median could differ in variability. Either they could differ in range (uniform) or shape. Differently spread bell-shaped histograms of common mean, median, and range are natural fodder for investigation. Students can manipulate the data so that a peak in the middle is achieved while the tails become narrow. The goal of such an activity would be for students to be able to make valid comparisons between shape and relative variability.

To facilitate understanding the connection between measures of center and shape of distribution, instructors might consider first having students find measures of center from graphs, such as a stem-and-leaf plot, where the raw data is completely accessible. Students can discover connections regarding the relative positions of measures of center with respect to shape and then more easily make generalizations to similar types of more abstract graphs such as histograms of ungrouped and grouped data.