Effect of Prior Information on String Length Estimates

KEYWORDS:
Teaching;
Displaying data;
Describing distributions;
Comparing distributions;
Properties of measurements.

Christine A. Franklin
University of Georgia, Athens, Georgia, USA.
e-mail: chris@stat.uga.edu

Madhuri S. Mulekar
University of South Alabama, Mobile, Alabama, USA.
e-mail: mmulekar@jaguar1.usouthal.edu

Summary
This article describes an activity through which students collect data and explore ways to display them through graphs and charts. It also motivates various summary measures for location, spread and shape. Finally, it gives an introduction to concepts of validity, reliability and unbiasedness.

◆ INTRODUCTION ◆

Students often view collected data as a list of numbers and fail to recognize that when summarized, collectively, they provide useful information for decision making. The activity described in this paper gives students an opportunity to visualize data through various graphs and charts, and to explore and compare different shapes of distributions. It also allows them to discover reasons for changes in shapes, centres and spreads of distributions.

◆ BACKGROUND AND GOALS ◆

The authors of this paper used (and still use) this activity in a non-calculus-based introductory-level statistics course. It is also recommended for students taking a statistics course at the high-school level or for students in science courses planning to collect data. Even before they embark upon formal inferential or estimation procedures, students should learn about measurements, variation among measurements and properties of measurements, by exploring and summarizing collected data.

In addition to the general goals described above, we also hope that students will learn to:

- deal with large data sets
- explore data using distributions
- describe distributions using shape, centre and spread
- identify any unusual characteristics of the distributions, such as outliers or gaps
- explore three properties of measurements: validity, reliability and unbiasedness
- communicate results orally and in writing.

Riddiough and McColl (1998) described one activity related to the learning effect on repeated estimates, where they suggested obtaining repeated estimates of length of a string when no or some feedback is provided to students about the estimates on previous attempts. Scheaffer et al. (1998) presented a data collection and analysis activity using two strings. Here we present one activity that is an extension and enhancement of their activity. We also describe our experiences in using this activity in the classroom.

◆ COLLECTING THE DATA ◆

In this experiment, we used two strings of lengths 44 and 39 inches respectively. We referred to the 44-inch string as ‘String 1’ and the 39-inch string as ‘String 2’. During the opening remarks in class, we kept both of the strings wound up and did not unwind the two strings simultaneously. This prevented students from comparing string lengths, but assured them that there really were two different strings.

- We told students that we would show them two different strings and that they would be asked to estimate their lengths visually. No measuring tapes would be allowed. In other words, the students' eyes would be their only instrument available to take measurements.
- First we showed students String 1, holding it horizontally, pulling it tight from one end to the other, and asked them to write down to the nearest inch their estimate of the length.
- Then we warned the students against changing their measurements (guesses) for the first string.
- After giving this warning, we told students that we wanted to help them with their measurements and that the second string was at least 35 inches long.
- Then we showed students String 2. Again holding horizontally, pulling it tight from one end to the other, we asked them to write down to the nearest inch their estimate of the length.
- We gathered all the data and input the data into a spreadsheet for further analysis.

We collected data from four different classes using the described procedure and combined the measurements from the four classes into one data set. This data set of string lengths can be accessed from http://www.stat.uga.edu/faculty/FRANKLIN/strings.XLS. The data set contains lengths estimated for Strings 1 and 2. It also identifies the classes from which the data were collected.

Since we taught the classes in a computer lab, our students had access to computers for exploring the data set which they had helped to create. If students do not have access to such computers, then the instructor may create graphical and numerical summaries and provide students with copies.

Before showing graphical displays, we initiated a class discussion about the characteristics of distributions for the estimated lengths. Starting with String 1, we asked students to predict the shape of the expected distribution of lengths for each string and the reason for predicting this shape. For String 1, they expected some very low estimates and some very high estimates, but that most estimates would cluster in the middle, yielding a fairly *symmetric* distribution. The students predicted that the String 2 distribution would be truncated at 35 with a longer right tail. They recognized that, due to the knowledge of a lower limit on the length of String 2, there wouldn't be any extremely low estimates, but that most of the estimates would be from 35 inches to around 40 inches with only a few students estimating the string to be longer. This would make the distribution *right skewed*, i.e. a distribution with a mound on the lower end and a longer right tail. Of course, we did not reveal actual lengths or the distribution characteristics at this point in our discussion, and moved on with the analysis of the data set.

◆ RESULTS AND DISCUSSION ◆

Note that all graphs and numerical summaries presented below were created using the statistical software Minitab. However, any statistical software would work well with this activity.

String 1 results

In response to students' suggestions, we started with a histogram and a dotplot of the data for String 1 (figures 1 and 2). Students quickly realized that the distribution of length estimates for String 1 was fairly symmetric, as they expected. Some considered it to be slightly right skewed. We

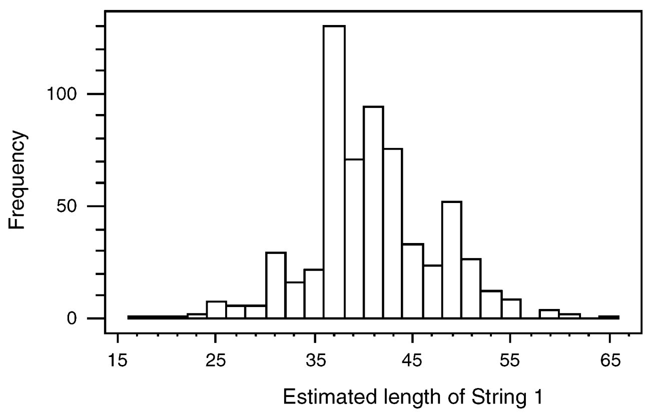

Fig 1. Histogram of estimated length of String 1

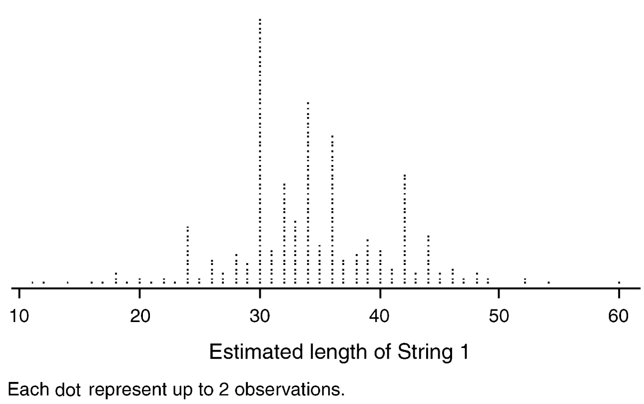

Fig 2. Dotplot of estimated length of String 1

Variable	*N*	*Mean*	*Median*	*TrMean*	*StDev*	*SE Mean*	*Minimum*	*Maximum*	*Q1*	*Q3*
String 1	622	40.079	40.000	40.064	6.559	0.263	17.000	66.000	36.000	44.000

Table 1. Numerical summary statistics for estimated length of String 1

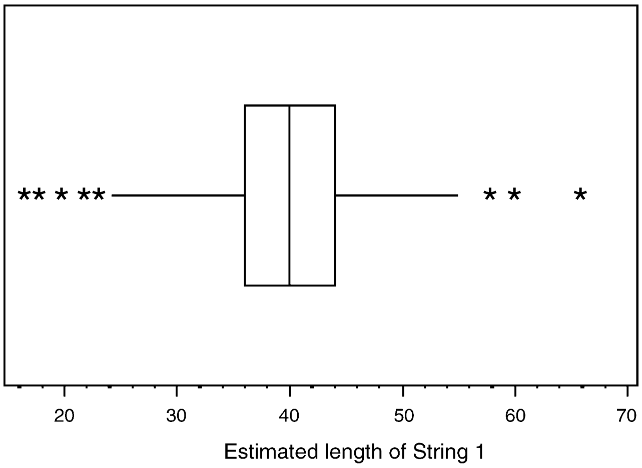

Fig 3. Boxplot of estimated length of String 1

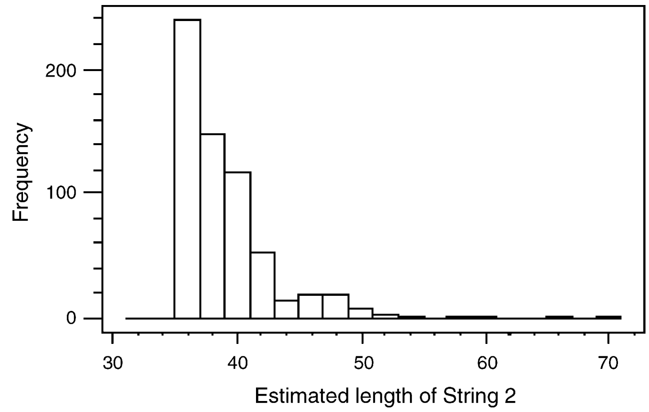

Fig 4. Histogram of estimated length of String 2

then discussed where the length estimates in the distribution would be centred. Letting each of the estimated length measurements be thought of as the location of identical weights strung out along a line, the point at which the line balanced was estimated to be approximately 40 inches. This balancing point is the *mean*. The range of measurements was noted to be approximately 50 inches.

Next, we generated a boxplot and a numerical summary (figure 3 and table 1).

The numerical summary (and boxplot) showed the median to be 40 inches. In other words, half the class guessed the length of String 1 to be 40 inches or lower, while the other half guessed it to be 40 inches or higher. The guesses ranged from 17 inches to 66 inches. The numerical summary also showed that the range of the middle 50% of measurements was 8 inches (36 to 44 inches). This range is known as the *interquartile range* (IQR) and is computed as the difference between the third and the first quartile ($IQR = Q_3 - Q_1$).

In discussing the boxplot, we reminded students that we would be using what is sometimes called a modified boxplot, instead of the straightforward box-and-whisker plot. A modified boxplot is useful in identifying possible outliers. In the box-and-whisker plot, lines (or whiskers) are drawn from the central box to the smallest and largest observations in the data set, with no indication that an observation(s) may be a potential outlier. All the boxplots that we constructed in this activity were modified boxplots.

Students described the boxplot for String 1 as approximately *symmetric*, i.e. the median centred in the central box and whiskers from the box extending to similar lengths. Using stars, the boxplot identified several outliers on both ends. The outliers were flagged based on the 1.5 * *IQR* rule. Any measurement outside the boundaries $Q_1 - 1.5 * IQR$ and $Q_3 + 1.5 * IQR$ was identified as an *outlier*, i.e. an observation not conforming to the general trend or pattern of the measurements. This led to the discussion of possible reasons for the outliers. We discovered that the largest measurement (at 66 inches) was from a non-American student who was more familiar with the metric system.

At this point, we informed students that the actual length of String 1 was 44 inches. They soon realized that about 75% of the students estimated it at or below 44 inches.

String 2 results

We then proceeded with a similar analysis of data for String 2. The histogram and the dotplot for the String 2 data were generated as shown in figures 4 and 5. Students identified the distribution to be extremely right skewed. Since the lower limit on the length of the string was given to them, they had expected all the measurements to cluster above

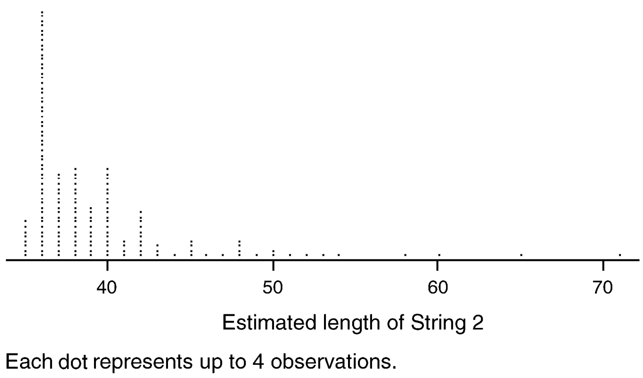

Fig 5. Dotplot of estimated length of String 2

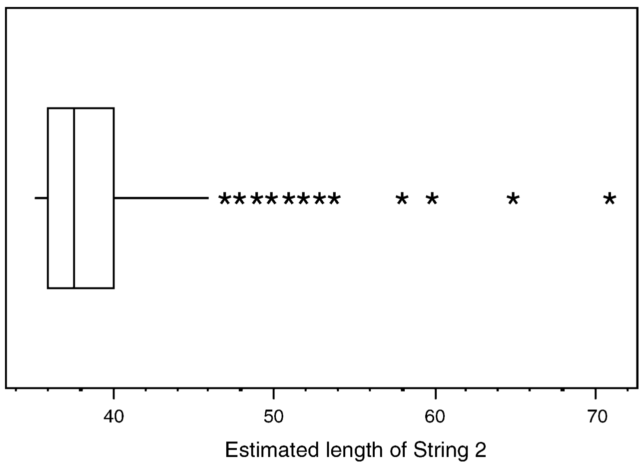

Fig 6. Boxplot of estimated length of String 2

this bound with a few extremely high measurements. Judging from the balancing point, the mean estimated length was about 39 inches. The range of the distribution was approximately 35 inches. This was less than the range for String 1 due to the absence of extremely low estimates.

Next we created a boxplot and numerical summary (figure 6 and table 2).

The numerical summary (and boxplot) showed the median to be 37.5 inches. In other words, half the class guessed the length of String 2 to be 37.5 inches or lower, while the other half guessed it to be 37.5 inches or higher. The guesses ranged from 35 inches to 71 inches. The range of the middle 50% of measurements, i.e. the IQR, was 4 inches (36 to 40 inches). The boxplot was also *right skewed*, i.e. the median pulled to the left of the central box and the right whisker longer than the left whisker. The boxplot also identified several outliers on the higher end. This led to a discussion of possible reasons for the outliers. Students immediately realized that, due to prior information available about the length of String 2, there were no extremely low guesses. Since the lower bound was given to them, most of the measurements clustered above this bound with a few extremely high estimates. The extremely high observation at 71 inches was given by the student more comfortable with the metric system.

The effect of the outliers on the mean and standard deviation was also noted by the students. Because of the extreme measurements, the mean was greater than the median. Given that 75% of the measurements ranged from 35 to 40 inches, the students also commented that the standard deviation of 4.065 inches was inflated for this distribution. They expected an average deviation from the mean to be approximately 2 inches, not 4 inches. Using this opportunity, we introduced the concept of a resistant numerical summary versus a non-resistant numerical summary. A *resistant numerical summary* is not affected by outliers. The median and IQR are resistant numerical summaries, whereas the mean and standard deviation are non-resistant summaries.

At this point, we informed students that the actual length of String 2 was 39 inches. They soon realized that about 75% of students estimated it at or below 39 inches.

◆ COMPARISON OF TWO ◆ DISTRIBUTIONS

In order to facilitate comparison of two distributions, we created parallel boxplots (figure 7).

The parallel boxplots clearly showed differences between the two distributions. Although the distribution of estimated lengths was more symmetric for String 1 and more right skewed for String 2, the mean estimated length for String 2 was much

Variable	*N*	*Mean*	*Median*	*TrMean*	*StDev*	*SE Mean*	*Minimum*	*Maximum*	*Q1*	*Q3*
String 2	622	38.748	37.500	38.245	4.065	0.163	35.000	71.000	36.000	40.000

Table 2. Numerical summary statistics for estimated length of String 2

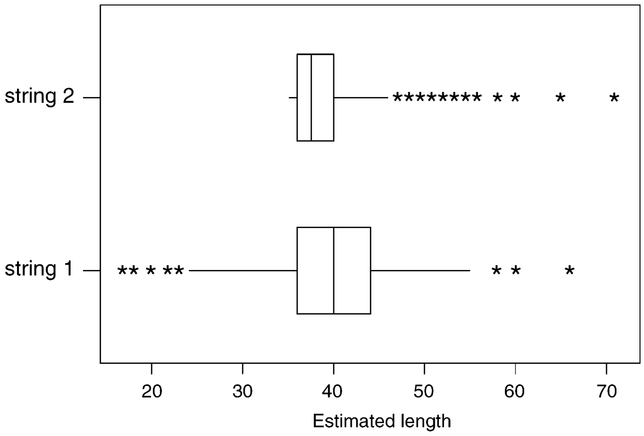

Fig 7. Parallel boxplots for estimated string lengths

closer to the true length than that for String 1. The spread of the estimated lengths was less for String 2, indicating that more of the estimated lengths were closer to the true length for String 2 than those for String 1. In other words, with the prior information about the lower bound on the length available, more students were able to estimate a string length closer in value to the actual length.

Some students commented at this point that if we had held the strings vertically, instead of horizontally, their estimates would have been closer to the actual length. In their opinion, it would have been easier to gauge the string length by using our body heights rather than our arm spans for comparison. We don't know if there is any evidence to support their belief. That could be the subject of another study.

◆ PROPERTIES OF MEASUREMENTS ◆

At this point in the discussion, some students raised questions about the appropriateness of combining measurements from four different classes and possible differences from class to class. We used this opportunity to introduce students to the process of obtaining measurements and three properties of measurements: validity, reliability and unbiasedness.

A measurement process is considered *valid* if it is appropriate for measuring a desired property; in our case, the length of the two strings. Is 'using the eyes' as the instrument for obtaining the measurements a valid way to measure the length of the strings? We believe that it is a common and acceptable method. Some students suggested that using eyes as an instrument was not a valid measurement process since measurements using eyes are generally less accurate than measurements obtained using a ruler/tape. Although students used the term 'validity', their reasoning showed that they actually meant 'accuracy' of the measurement process. So we steered the discussion to the difference between *accuracy* and *validity*.

The *accuracy* of the measurements can be analysed by two components or properties: reliability and unbiasedness. A measurement process is *reliable* if the measurements are repeatable. A measurement process is *unbiased* if the mean of the entire set of possible estimates is equal to the actual unknown value being estimated, in this activity the true length of the strings. In other words, a measurement process is *unbiased* if it doesn't systematically overestimate or underestimate the actual value of the desired characteristic being measured.

Keeping in mind the concern raised by students, we decided to create parallel boxplots for measurements from the four classes. The resulting boxplots are shown in figures 8 and 9.

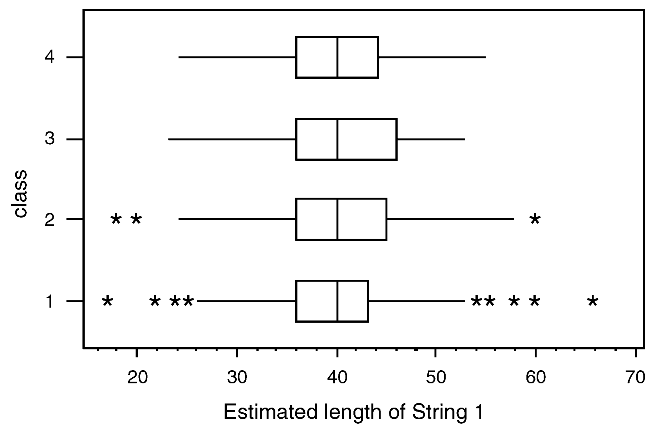

Fig 8. Estimated length of String 1 by class

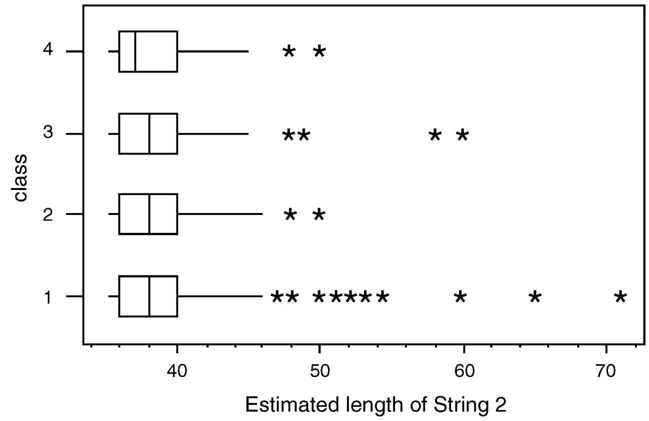

Fig 9. Estimated length of String 2 by class

	Class	N	Mean	Median	TrMean	StDev	SE Mean	Minimum	Maximum	Q1	Q3
String 1	1	321	39.900	40.000	39.837	6.553	0.366	17.000	66.000	36.000	43.000
	2	172	40.238	40.000	40.240	6.874	0.524	18.000	60.000	36.000	45.000
	3	67	40.403	40.000	40.475	6.472	0.791	23.000	53.000	36.000	46.000
	4	62	40.210	40.000	40.357	5.876	0.746	24.000	55.000	36.000	44.250
String 2	1	321	38.863	38.000	38.273	4.402	0.246	35.000	71.000	36.000	40.000
	2	172	38.593	38.000	38.240	3.370	0.257	35.000	50.000	36.000	40.000
	3	67	39.164	38.000	38.541	4.653	0.569	35.000	60.000	36.000	40.000
	4	62	38.129	37.000	37.732	3.247	0.412	35.000	50.000	36.000	40.000

Table 3. Summary statistics for string lengths by class

The parallel boxplots (figure 8) showed that all four classes had similar means, ranges, median estimated lengths and IQRs for String 1. Estimates from all four classes resulted in fairly symmetric distributions. It seemed that all four classes gave fairly similar estimates for the String 1 lengths, with the exception that classes 1 and 2 had a few outliers. Similarly, from parallel boxplots for String 2 (figure 9), students saw that distributions of estimated lengths from all four classes were right skewed. All of them had similar medians and IQRs. Both graphs showed how consistently students in all four classes estimated the string lengths. Students quickly realized that the process of measuring the strings using one's eyes was very *reliable* or repeatable from one class to the next.

How close were the measurements to the actual lengths of the strings? In search of an answer to this question, we computed the summary statistics for string lengths by class (see table 3). Remember that Strings 1 and 2 were of lengths 44 and 39 inches respectively. The String 1 average or mean estimated length for the distribution of measurements was about 4 inches less than the actual length of 44 inches. The String 2 average or mean estimated length for its distribution was about 0.1 to 0.9 inches below the actual length of 39 inches. The difference between the mean estimated length and the actual length is known as the *bias*. The summary statistics showed that all four classes consistently gave less biased estimates for lengths of String 2 than for lengths of String 1. Note that the variation in estimates of String 2 lengths was also less than that for String 1 lengths for all four classes. The students commented that it helped the estimation process to have some prior knowledge about the string length.

We noticed some misconception about the term 'bias'. Some students commented that one of their estimated string lengths was exactly the same as or very close to the actual value; thus, they claimed their measurement was not biased. We indicated that more measurements would be necessary to adequately judge if an individual's eyes were on average giving an unbiased measure. The classes' eyes were clearly on average giving measures that were biased. We noted that measurement bias is a product of the measurement process, not to be judged based on one measurement from that process. Our discussion and data analysis helped clarify this misconception about bias.

◆ CONCLUSIONS ◆

Gathering measurements on the lengths for the two strings allowed the creation of a student-generated data set that students could use to explore data by choosing appropriate graphical and numerical techniques. The string data also helped in exploring three important properties of measurement: validity, reliability and unbiasedness. This activity encouraged students to communicate the behaviour of the string length data, both orally and in writing. Consistent oral and written communication helped the students give clear, concise and non-technical descriptions of the data, both statistically and in the context of the scenario being analysed.

References

Riddiough, R. and McColl, J.H. (1998). How long is a piece of string? *Teaching Statistics*, **20**(1), 13–16.

Scheaffer, R., Gnanadesikan, M., Watkins, A. and Witmer, J. (1996). *Activity-Based Statistics*. New York: Springer.