

# Step Into Statistics

Subhash Bagui  
University of West Florida  
[sbagui@uwf.edu](mailto:sbagui@uwf.edu)



**Published: March 2012**

## Overview of Lesson

In this activity, students will collect data in their class to determine foot sizes of students and compare foot sizes between boys and girls. To illustrate the data students will compute measures of center (mean, median), quartiles (first, second, and third quartiles) and spread (range, Interquartile Range (IQR), standard deviation) for boys and girls combined and also separately. Students will also construct a comparative boxplot to compare foot sizes of boys and girls. Conclusions will be drawn based upon analysis of the data and examinations of graphs in the context of questions asked about sizes of feet.

## GAISE Components

This investigation follows the four components of statistical problem solving put forth in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report. The four components are: formulate a question, design and implement a plan to collect data, analyze the data by measures and graphs, and interpret the results in the context of the original question. This is a GAISE Level B activity.

## Common Core State Standards for Mathematical Practice

1. Make sense of problems and persevere in solving them.
2. Reason abstractly and quantitatively.
3. Construct viable arguments and critique the reasoning of others.
4. Model with mathematics.
5. Use appropriate tools strategically.

## Common Core State Standards for Mathematical Practice (Grade 7)

7. SP. 3. Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.
7. SP. 4. Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.

## NCTM Principles and Standards for School Mathematics

### Data Analysis and Probability Standards for Grades 6-8

**Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them:**

- formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population;
- select, create, and use appropriate graphical representations of data, including histograms, box plots, and scatterplots.

### **Select and use appropriate statistical methods to analyze data:**

- discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots;
- find, use, and interpret measures of center and spread, including mean and interquartile range.

### **Prerequisites**

Students should have prior knowledge of making measurements (in cm) using a ruler, organizing data in a table, calculating descriptive statistics, creating dotplots and boxplots, and how to interpret descriptive statistics.

### **Learning Targets**

Students will be able to construct a comparative boxplot that compares two groups and a separate boxplot for combined data. In addition, students will be able to analyze, compare, and interpret the (statistical) analysis of the data.

### **Time Required**

Approximately one 45-50 minute class period; some discussion time the following class period might be necessary.

### **Materials Required**

Graphing calculator or computer with spreadsheet or statistical software that can be used to create and print boxplots, rulers with centimeters, data recording sheet, poster board to display the data.

### **Instructional Lesson Plan**

#### **The GAISE Statistical Problem-Solving Procedure**

##### **I. Formulate Question(s)**

Begin the lesson by telling the class that it is of interest to examine foot sizes of people. More specifically, a statistical study can be conducted to study foot sizes of students in the class. This information can be very important to a shoe manufacturer. Information obtained from these types of investigations can help companies determine what sizes of shoes they should manufacture for middle school students.

Ask students to write some questions that they would be interested in investigating about students' foot lengths. Some possible questions might be:

1. What is the mean (representative) foot size of the class? What is the median (typical) foot size of the class? What is the shortest foot size in the class? What is the longest foot size in the class?
2. Are there differences in foot sizes for boys and girls? If so, what are the differences?
3. Are foot sizes related to any other variables?

##### **II. Design and Implement a Plan to Collect the Data**

In the data collection phase ask students what kind of measurements should be made? Make sure that students talk about how to make the measurements accurately and with precision. Ask the

class who should measure the students' feet. More than one person might be helpful in the data collection phase. It might be a good idea for one person to measure all students and for another to record their foot sizes and gender. Before collecting data ask students to decide which foot should be measured? Also tell students to round measurements to the nearest  $\frac{1}{2}$  centimeter. This way a protocol for measurements can be developed and it will ensure consistency from measurement to measurement.

Measure all students and record their foot sizes (in cm) and gender in a data table. A sample class data set is shown in the table below; a blank data table is provided on the Activity Sheet on page 10.

Table 1. Sample class data.

Student	Gender	Foot Length (cm)
1	F	22.5
2	F	23.0
3	F	23.5
4	F	24.0
5	F	24.5
6	F	24.5
7	F	25.0
8	F	25.0
9	F	26.0
10	F	26.0
11	F	26.0
12	M	25.5
13	M	25.5
14	M	26.0
15	M	27.0
16	M	27.0
17	M	27.5
18	M	28.0
19	M	28.5
20	M	29.0

Ask the students to explain why this is an observational study and not an experimental study. Tell students that data values are recorded from direct observation and measurements. Nothing has been done deliberately to the students in order to collect data.

### III. Analyze the Data

Different statistical tools are used for analysis of different questions. For example, the class can calculate the mean and spread from the collected data. A boxplot can be constructed from the same data set. The same analyses can be repeated separately for the boys and girls in class. Ask

students to suggest graphs that might be useful to compare the foot length data distributions for boys and girls. Comparative boxplots are appropriate for displaying this data.

To analyze the foot sizes, the class can calculate measures of center and spread and create boxplots. To create a boxplot one needs the 5-number summary: minimum, first quartile (25<sup>th</sup> percentile,  $Q_1$ ), second quartile (median, 50<sup>th</sup> percentile,  $Q_2$ ), third quartile (75<sup>th</sup> percentile,  $Q_3$ ), and maximum. First these five numbers are plotted on a line extended from the minimum to the maximum and then a box is created around  $Q_1$  and  $Q_3$  with lines drawn at the first quartile, the second quartile, and the third quartile. The difference between the third and first quartiles is called the interquartile range (IQR).

Descriptive Statistics for the 20 sample class foot sizes are calculated. The numerical calculations show that the mean foot size of the class is 25.70 cm. The median foot size is 25.75 cm. The teacher can discuss with the class that the median represents the 50<sup>th</sup> percentile of the distribution of the class foot sizes. About half of the students have foot sizes less than 25.75 cm and another half of the students have foot sizes more than 25.75 cm. The shortest foot size in the class is 22.50 cm; the longest foot size is 29.00 cm. About one fourth (25%) of the students have foot sizes below 24.50 cm (the first quartile) and one fourth (25%) of the students have foot sizes above 27.00 cm (the third quartile). About half (50%) of the students have foot sizes between 24.50 and 27.00 cm. The standard deviation (sd) of 1.78 cm provides a typical difference between the student foot sizes and the mean.

The following rule is used to find an outlier (extreme value) in a data set. A foot size ( $X$ ) is called an outlier if  $X > Q_3 + 1.5(IQR)$  or  $X < Q_1 - 1.5(IQR)$ . For the combined (boys and girls) data set  $Q_1 = 24.5$ ,  $Q_3 = 27.0$ , and  $IQR = Q_3 - Q_1 = 27.00 - 24.50 = 2.5$ . So  $1.5(IQR) = (1.5)(2.5) = 3.75$ . Thus,  $Q_3 + 1.5(IQR) = 27.00 + 3.75 = 30.75$  and  $Q_1 - 1.5(IQR) = 24.5 - 3.75 = 20.75$ . Thus, any foot size value ( $X$ ) in the combined data greater than 30.75 cm or smaller than 20.75 cm would be an outlier. Similar rules can be formulated for boys and girls separately to find outliers in the respective groups.

The descriptive statistics for the sample boys' data are given by:  $n=9$ , Mean=27.11, Standard Deviation=1.27, Min=25.50, First Quartile =25.75, Median =27.00, Third Quartile=28.25, and Maximum=29.00. These descriptive statistic values can be interpreted similar to above.

Similarly, the descriptive statistics for the sample girls' data are given by:  $n=11$ , Mean=24.55, Standard Deviation=1.21, Min=22.50, First Quartile =23.50, Median =24.50, Third Quartile=26.00, and Maximum=26.00. These descriptive statistic values can also be interpreted similar to above.

A boxplot for the class foot sizes is shown in Figure 1. The boxplot depicts the 5-number summary of the class heights. The plot shows that the median foot size of the class is 25.75 cm. The middle 50% of the class foot sizes ranges from 24.50 cm to 27.00 cm (as seen by IQR). Thus here  $IQR = 27.00 - 24.50 = 2.50$  cm. The shortest foot size is about 22.50 cm and the longest foot size is about 29.00 cm. Similarly, a comparative boxplot can be created for boys and girls. Students may use appropriate technology (graphing calculator, Excel, statistical software) to create these plots. The comparative boxplot is shown in Figure 2 below.

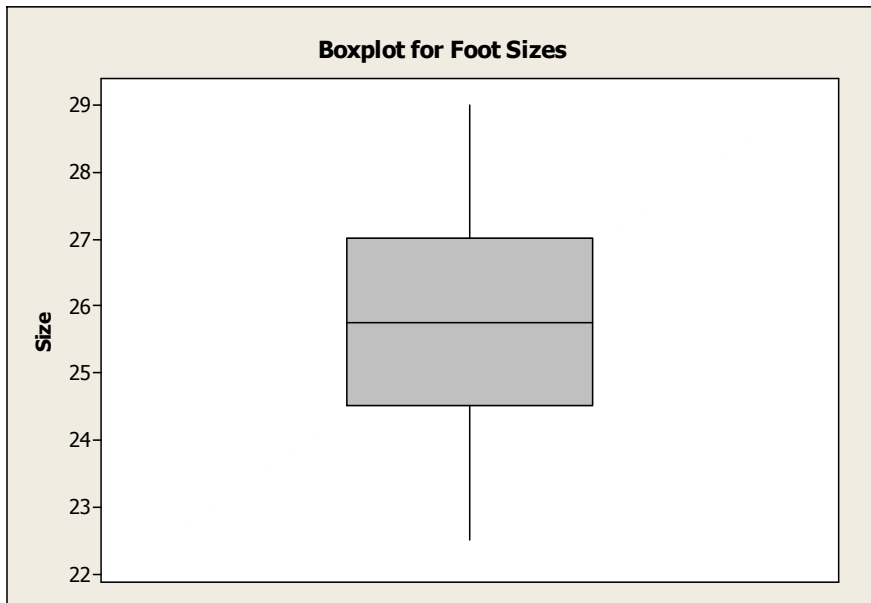


Figure 1. Boxplot for sample class foot sizes.

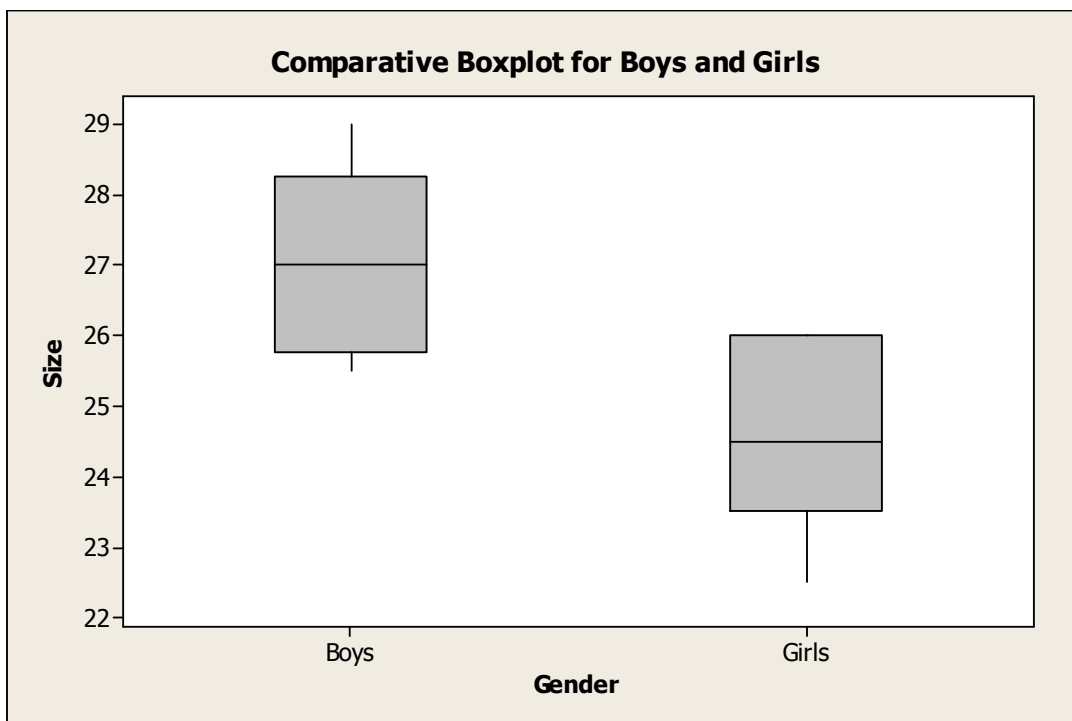


Figure 2. Comparative boxplot for boys and girls.

Have students view the above comparative boxplot for boys and girls. Ask the class if there appears to be any evidence of boys or girls having a higher median foot size. Ask the class if the genders show similar variability in foot sizes.

#### IV. Interpret the Results

From Figure 1 we notice that the minimum value (22.5), first quartile (24.5), third quartile (27.0), and maximum value (29.0) are symmetrically located around median (25.75). This indicates that the class foot sizes may be symmetrically distributed.

From the comparative boxplot (Figure 2) we notice that boys tend to have bigger foot sizes than girls. Also, boys have a much higher median foot size than girls. All foot sizes for boys are longer than at least 75% of the foot sizes for girls.

The next step in the analysis is to focus on measures of center. The mean and median foot lengths for boys are 27.11 and 27.00 cm, respectively. For girls, the mean and median foot lengths are 24.55 and 24.50 cm, respectively. As expected, it seems that a typical boy's foot length is bigger than a typical girl's foot length by about 2.5 cm.

Next ask students to characterize the spread of the foot length distributions. The spread is generally measured using range (max-min), standard deviation (sd), or IQR. The spread summaries for boys and girls are shown in the table below:

	Range	SD	IQR
Boys	3.5	1.27	2.5
Girls	3.5	1.21	2.5

The spread for boys and girls with respect to range, standard deviation, and IQR are very comparable. Ask students to determine if there are any outlying values. Ask students if foot sizes are related to any other variables? Ask students if they would feel comfortable generalizing these results to the population of students at the school?

## Assessment

1. A random sample of 20 students is selected and their gender and foot sizes in cm are recorded. The data are given in the following table.

Data table:

<b>Student</b>	<b>Gender</b>	<b>Foot Size (cm)</b>
1	Female	24
2	Male	25
3	Female	26
4	Male	24
5	Male	23
6	Female	23
7	Male	26
8	Female	22
9	Male	29
10	Male	27
11	Female	24
12	Male	26
13	Male	28
14	Male	23
15	Female	24
16	Male	27
17	Male	25
18	Female	23
19	Male	27
20	Male	27

Using the above data answer the following questions:

- (a) Calculate the 5-number summary to create a boxplot for the foot sizes of students.

(b) Construct the boxplot for foot sizes.

(c) What is the mean foot size? What is the median foot size? What are the shortest and longest foot sizes? Give two numbers that cover the middle 50% of the distribution of the foot sizes. What is the range of the most common foot sizes?

(d) What is the mean foot size for boys? What is the median foot size for boys? What are the shortest and longest foot sizes for boys? Give two numbers that cover the middle 50% of the distribution of the foot sizes of the boys. What is the range of the most common foot sizes for boys?

(e) What is the mean foot size for girls? What is the typical foot size for girls? What are the shortest and longest foot sizes for girls? Give two numbers that cover the middle 50% of the distribution of the foot sizes of the girls. What is the range of the most common foot sizes for girls?

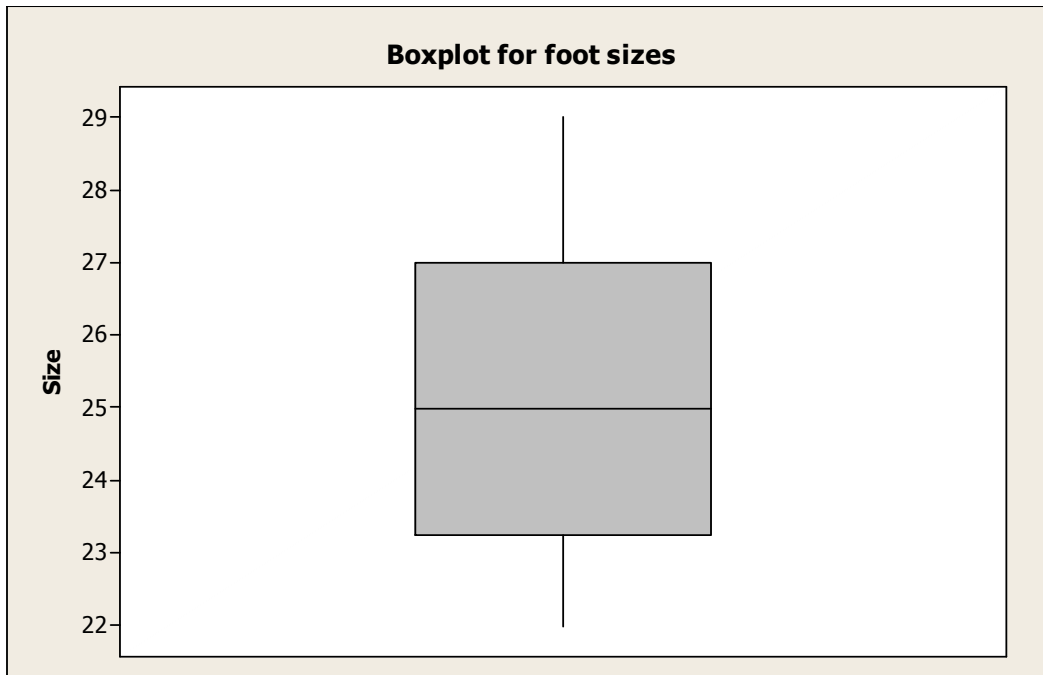


- (f) Using appropriate technology (graphing calculator, Excel, statistical software) construct a comparative boxplot. Do boys or girls generally have bigger feet? Is there any outlier either of boys or girls?

## Answers

(a) 5-number summary: minimum=22.00, first quartile=23.50, median=25.00, third quartile=27.00, and maximum=29.00.

(b) See the boxplot for foot sizes below:

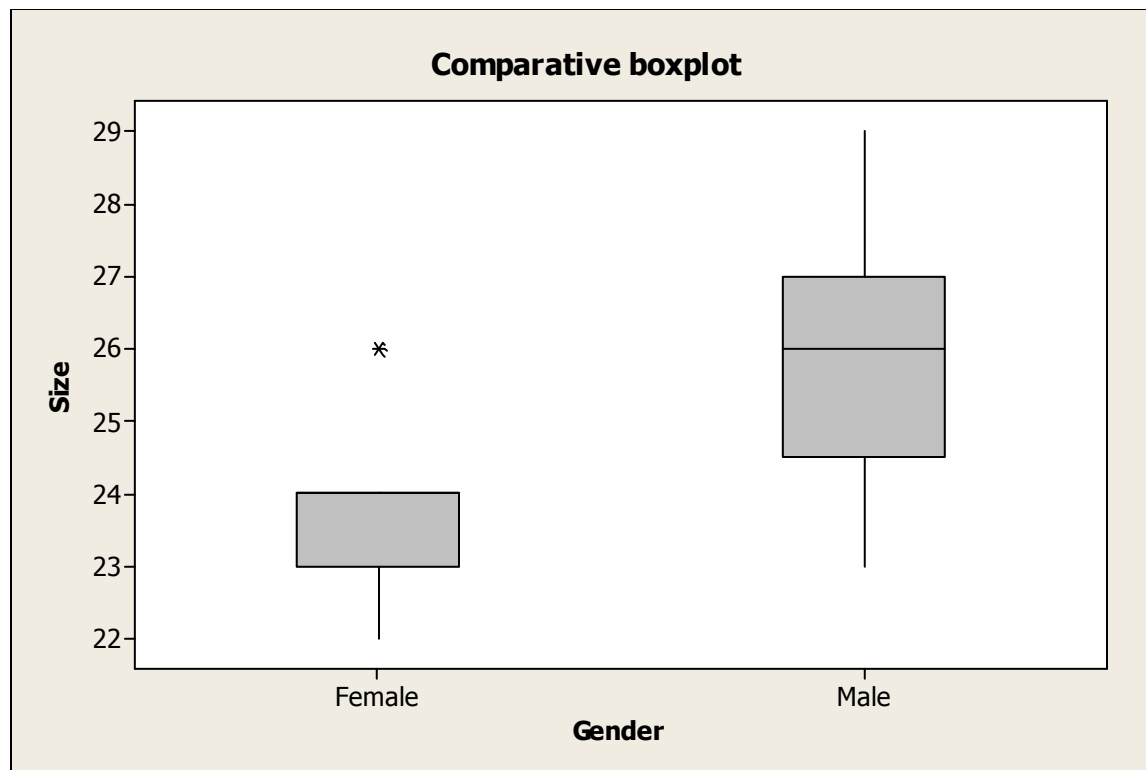


(c) Mean foot size= 25.15, median foot size=25.00, shortest foot size=22.00, longest foot size=29.00. The middle 50% of the distribution of the foot sizes are between 23.50 cm and 27.00 cm. Most common foot sizes range from 23.50 to 27.00 cm.

(d) Mean foot size for boys= 25.92, median foot size for boys=26.00, shortest foot size for boys=23.00, longest foot size for boys=29.00. The middle 50% of the distribution of the foot sizes for boys are between 24.50 cm and 27.00 cm. Most common foot sizes for boys range from 24.50 to 27.00 cm.

(e) Mean foot size for girls= 23.71, median foot size for girls =24.00, shortest foot size for girls =22.00, longest foot size for girls=26.00. The middle 50% of the distribution of the foot sizes for girls are between 23.00 cm and 24.00 cm. Most common foot sizes for girls range from 23.00 to 24.00 cm.

(f) See the comparative boxplot below. The boxplots show that boys generally have longer foot sizes than the girls' foot sizes. The median foot size for males is much higher than the median foot size for females. There is one high outlier in the female group.



- In the above plot \* shows an outlier (extreme value) in the foot sizes for girls. Note that there is a foot size of 26 cm of a girl in the sample.

### Possible Extensions

Intuition tells us that foot size can be a good predictor for height. So an extension of this experiment is to see if there is any relationship between foot size and height of students. Start by asking the class that if they expect the relationship between height and foot size to be linear.

### References

1. Adapted from an activity created by Paul J. Fields, Ph.D. for the *American Statistical Association Meeting Within a Meeting Program for Middle School Teachers* (2008).
2. *Sampling in Archaeology* by Mary Richardson 2012, <http://www.amstat.org/education/STEW/>.

### Step Into Statistics Activity Sheet

In this activity, you will collect data to assess foot sizes of students in your class and compare foot sizes between boys and girls.

1. Describe the data collection process that will be used to answer the research question.

2. Record gender and foot size for you and your classmates in the following table:

Student	Gender	Foot Size (cm)

3. Sort the foot sizes of the students from smallest to largest.

4. Using the sorted data, find the values needed to create a boxplot of the data.

**5.** Construct the boxplot for foot sizes:

**6.** What is the mean foot size of the class? What is the median foot size of the class? What is the shortest foot size in the class? What is the longest foot size in the class? Give two numbers that cover the middle 50% of the distribution of the foot sizes. What is the range of the most common foot sizes for the class?

**7.** Sort the foot sizes of boys from smallest to largest.

**8.** Using the sorted data, find the values needed to create a boxplot of the data for boys.

**9.** Sort the foot sizes of girls from smallest to largest.

**10.** Using the sorted data, find the values needed to create a boxplot of the data for girls.

**11.** What is the mean foot size for boys? What is the median foot size for boys? What are the shortest and longest foot sizes for boys? Give two numbers that cover the middle 50% of the distribution of the foot sizes of the boys. What is the range of the most common foot sizes for boys?

**12.** What is the mean foot size for girls? What is the typical foot size for girls? What are the shortest and longest foot sizes for girls? Give two numbers that cover the middle 50% of the distribution of the foot sizes of the girls. What is the range of the most common foot sizes for girls?

**13.** Using appropriate technology (graphing calculator, Excel, statistical software) or by hand construct a comparative boxplot. Do boys or girls generally have bigger feet? Are there any outliers for the boys or girls?