**Unit 5: Investigation 1 (3 Days)**

**One Variable Data**

***CCSS:*** S-ID 1; S-ID 2; S-ID 3

**Overview**

Students will explore and define measures of center and measures of spread and display data in dot plots, histograms, and box-and-whisker plots. Students will analyze real world data by hand and using a graphing calculator.

**Assessment Activities**

**Evidence of Success: What Will Students Be Able to Do?**

Students will be able to find and understand measures of center as well as measures of spread. They will be able to create and interpret a dot plot, histogram and box-and-whisker plot.

**Assessment Strategies: How Will They Show What They Know?**

**Exit Slip 5.1.1** asks students to construct a frequency table and histogram.

**Exit Slip 5.1.2** asks students to find the five-number summary, range, and interquartile range and construct a box-and-whisker plot.

**Journal Entry** asks students to describe the difference between the mean and the median and how changing the data set affects each measure of center.

**Launch Notes**

Begin the investigation by discussing Hurricane Isaac, which hit the city of New Orleans in August 2012 on the seventh anniversary of Hurricane Katrina. You may want to show a video of the destructive power of the hurricane such as the one at [www.**youtube.com**/watch?v=uYZzrgRm4nw](http://www.youtube.com/watch?v=uYZzrgRm4nw).

**Closure Notes**

Activity 5.1.7 ties together the all of the concepts developed in this investigation and is designed to be done in groups followed by a whole class discussion. As you wrap up the discussion, ask the class what have they have learned about measures of center and measures of spread and how histograms and box-and-whisker plots convey information about a set of data.

**Teaching Strategies**

1. 2005 was a particularly bad year for hurricanes in the Atlantic Ocean. Introduce the data table in **Activity 5.1.1** **Hurricanes**. Review the definitions of mean, median, and mode, which students should be familiar with from previous courses. As students compute these statistics for the hurricane data they will find that three numbers (75, 80, 85) are tied for the mode. That is why mean and median are more frequently used as measures of center.

Hurricanes are classified by category according to their maximum wind speed. Students will classify each of the fifteen hurricanes of 2005 by category and calculate the measures of center on the category variable. In this data set there is an unambiguous mode for category 1 hurricanes, which appears as the tallest bar when a dot plot is constructed.

At this point you may introduce students to the graphing calculator. The hurricane data are entered into L1 under the STAT menu. Students will then use the calculator to compute “one variable statistics.” From the maximum and minimum values they will compute the range.

Tell the students to save the data in their calculator as the data will be used later to create a histogram. Students first create a histogram by hand in questions 14 and 15 in the **Activity 5.1.1**. Then in questions 16 and 17 they create a histogram on the calculator. One advantage of using the calculator is that the width of the bins may be easily changed. Students may observe the effect of changing bin width on the number of and height of the bars.

As an extension, some students may perform a similar analysis for the 2012 hurricane season (pages 7 and 8 of **Activity 5.1.1**).

**Differentiated Instruction (Enrichment)**

Assign pages 7 and 8 of **Activity 5.1.1**. The instructions are much less structured than the previous ones. Question 22 asks them how the statistics and data displays can be used to draw conclusions about the difference between the two hurricane seasons.

 **Activity 5.1.2 Home Run Hitters** provides additional practice with measures of center, the range, and histograms. You may want to assign the first two pages as homework and save the third page, which involves looking at the shape of histograms for class discussion the next day. When students compare the homeruns in the American League and National League they see that the American League data is somewhat skewed with a tail to the right and the National League data is closer to being mound shaped.

 **Activity 5.1.3 More Histograms** may be used at any time during this investigation to give students more practice constructing and analyzing histograms. The one new skill introduced here is to use the Trace feature to read the frequency distribution from the display of the histogram in the Graph window.

**Differentiated Instruction (For Learners Needing More Help)**

To lessen the load, you can provide data lists that are already sorted from least to greatest. The lists can also be minimized so there are fewer numbers in the tables for the students to work with. Many students may find multiple columns difficult to deal with so one long column may provide more structure.

If students need help sorting data sets, they may try writing numbers on slips of paper and moving them around until they are in order.

**Exit Slip 5.1.1** assesses students’ ability to construct a frequency table and histogram. It may be given at the end of the first or second day of the investigation.

**Journal Entry**

Describe the difference between the median number and the mean number of a data set. What would happen to each measure if you added more numbers to your data set? What if you took away numbers? Make sure to include when each one would be most useful. You may use one of the examples you have completed to support you answer.

1. Discuss the difference between measures of center and measures of spread. So far students have learned three measures of center (mean, median and mode) and one measure of spread (range). Discuss the advantages and disadvantages of using the range to characterize the spread of a data set. On the one hand, knowing the range helps determine the scale when you are making a histogram since all of the data lies between the minimum and maximum. On the other hand, the range can give a distorted picture of how spread out most of the data is. You may use the example of gasoline prices from **Activity 5.1.3** to illustrate this. The range is almost a dollar ($.979). Yet it is strongly influenced by the maximum value (Hawaii, $4.605), which is $0.245 above the next highest value (Alaska, $4.360). For this reason statisticians look for other ways to measure spread. The two most common measures are the standard deviation (SD) and the interquartile Range (IQR).

You may show students that the calculator will find the standard deviation with the 1-Var Stats command. It actually shows both the population and sample standard deviations but for our purposes we will use the sample standard deviation designated by Sx. In the case of the gasoline price data, Sx = $0.182. Tell students they will learn more about standard deviation in another course; for now they should simply recognize it as a measure of spread.

The new measure of spread we are now going to focus on is the IQR. You may introduce IQR through a class exercise. Have the class line up in order according to some criterion. If you don’t think students will be embarrassed, an obvious criterion is height. Or if you prefer, assign each student a random number by drawing slips of paper out of a hat.

Find the median of the data set by having students step forward in pairs beginning with the two at the ends and moving inward until only one or two students are left in the original row. If it is one student, that student’s number is the median. Otherwise it is the average of the two numbers in the middle.

Now find the quartiles by finding Q1, the median of the group that lies below the median of the entire set, and Q3the median of the group that lies above the median of the entire set. Explain that the median and the quartiles divide the entire set into four (approximately) equal groups. The median may be thought of as Q2. The interquartile range is the difference Q3 – Q1. Have all students whose numbers lie between Q1 and Q3 step forward. The interquartile range shows the spread over the middle half of the data set.

The discussion should lead into **Activity 5.1.4 The Five-Number Summary**,where students find Q1 and Q3. In later activities they may find these statistics on the calculator using the command 1-Var Stats and scrolling down. Remind students that all data values must be ordered in order to determine the five-number summary by hand. You may want to do one or two examples from this activity in class and assign the rest for homework.

Students are now ready for **Activity 5.1.5 Outliers and the 1.5** $×$ **IQR Rule.** In this activity the concept of outlier is developed. Although there is no formal definition for outlier, it is a member of a data set that does not seem to fit the prevailing pattern. In the case of the test scores in the first example, the score of 33 is an obvious outlier, and if this score belongs to the student who missed four days of class before the test, we have an explanation why this score differs so much from the others.

One method statisticians use to identify outliers is the $1.5 × IQR$ rule. Outliers are defined as values that are more than 1.5 times IQR units below Q1 or above Q3. Show students how to compute the lower fence and the upper fence using the $1.5 × IQR$ rule and then determine whether a point in the data set is an outlier.

Before doing questions 2 and 3 in **Activity 5.1.5** you may want to discuss the controversy that arose in 2010 when basketball star LeBron James left the Cleveland Cavaliers to join the Miami Heat where he could team up with two other superstars, Dwayne Wade and Chris Bosh. How did James’s performance compare with those of his other teammates on the Cavaliers? In other words, was he an outlier? The $1.5 × IQR$ rule can help answer this question.

1. Once students are comfortable finding and interpreting the five-number summary, they are ready to display the five numbers in a special graph called a box-and-whisker plot. **Activity 5.1.6 Box-and-Whisker Plots** introduces students to this type of graph. The box shows the two quartiles and the median. Its length is equal to the IQR. The whiskers connect the edges of the box to the extreme values (minimum and maximum).

Students first construct the plot by hand and then with the calculator. One variation of the box-and-whisker plot allows the whiskers to extend only to data points that lie within the lower and upper fences. In this display outliers are indicate as discrete points, with special symbols such as an asterisk or small square or cross on the TI graphing calculators.

There are two versions of Activity 5.1.6. The questions are the same but the order is different. In Activity 5.1.6a students examine several data sets and make box-and-whisker plots by hand before using the calculator to make the plots. Toward the end of the activity they return to data sets used earlier. In Activity 5.1.6b students work with one data set at a time and answer all questions pertaining to that data set before moving on to another one. Use whichever version you prefer.

At this point you can compare and contrast histograms with box-and-whisker plots. Ask students as in question 8(e) how to identify an outlier from a histogram. (In the case of *Thriller* it will appear in a bar all by itself at one end of the distribution.)

**Exit Slip 5.1.2** assesses students’ ability to find the five-number summary and construct a box-and-whisker plot.

**Activity 5.1.7 Test Grades** is designed as a culminating activity for this investigation. It is particularly suited for group work as described below.

**Group Work**

**Activity 5.1.7 Test Grades** may be given to students to work on in groups. Students can then present their findings to the entire class. Since the different groups will be analyzing different data sets, the class as a whole may then compare and contrast the results from each group.

**Resources and Materials**

* **Activity 5.1.1** Hurricanes
* **Activity 5.1.2** Homerun Hitters
* **Activity 5.1.3** More Histograms
* **Activity 5.1.4** Five-Number summary
* **Activity 5.1.5** Outliers and the 1.5 $×$ IQR Rule
* **Activity 5.1.6** Box-and-Whisker Plots
* **Activity 5.1.7** Test Grades
* **Exit Slip 5.1.1** Marine Animals
* **Exit Slip 5.1.2** Calories in Fruit
* Graphing Calculators
* Bulletin board for key concepts
* Student Journals