

About the Lesson

In this activity, students create and explore a box plot and histogram for a data set. They then compare the two data displays by viewing them together and use the comparison to draw conclusions about the data. As a result, students will:

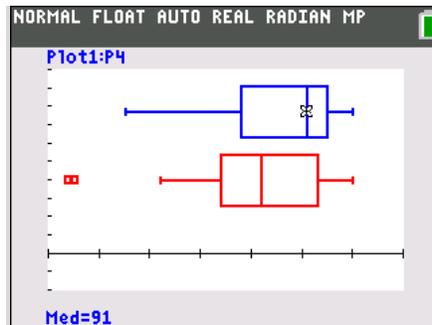
- Represent real-world data using a histogram and box plot.
- Calculate measures of central tendency and spread.
- Compare distributions from two data sets and draw conclusions.
- Identify and analyze outliers in a data set.

Vocabulary

- | | |
|-----------|-----------------------|
| • mean | • histogram |
| • median | • minimum |
| • skew | • interquartile range |
| • outlier | • box plot |

Teacher Preparation and Notes

- In this activity, students will compare two distributions. Graphical and numerical representations will be used to draw conclusions. Students will graph histograms to make comparisons based on shape, center, and spread. They will graph box plots on the same axes comparing shape, spread, and the five-number summary.
- Load list files P4, P5 on student graphing calculators.
- The activity is created so students can work independently or in small groups. Since students may struggle drawing conclusions and justifying them, it may be necessary to discuss valid conclusions towards the end of the activity.



Tech Tips:

- This activity includes screen captures taken from the TI-84 Plus C Silver Edition. It is also appropriate for use with the TI-84 Plus family with the latest TI-84 Plus operating system (2.55MP) featuring MathPrint™ functionality. Slight variations to these directions given within may be required if using other calculator models.
- Access free tutorials at <http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials>
- Any required calculator files can be distributed to students via handheld-to-handheld transfer.

Compatible Devices:

- TI-84 Plus Family
- TI-84 Plus C Silver Edition

Associated Materials:

- BestClass_Student.pdf
- BestClass_Student.doc
- P4 and P5.8xl (lists)

Problem 1 – Period 4 vs. Period 5 Test Results
Comparing distributions without graphing

Students are given a table of data showing the test results of two different class periods. The teacher wants to graphically display this data so the students can get a quick understanding for how the class did and how they compare with the other class period. Additional benefits of displaying the data graphically are that the students won't see individual results and they will refresh their understanding of histograms and box plots.

The answers to the first questions are expected to be only approximations since it is difficult to analyze data and draw conclusions from a table of values. Although it is possible for students to enter the data, it is highly recommended that they transfer the lists P4 and P5.

1. Without graphing the data, describe each data set. To better estimate the center and spread of each data set, sort using **STAT 3:SortD(**. Then use **[2nd] [STAT]** to select the list P4. Repeat for P5, period 5.

Answer: Students' answers should not be precise at this point. Students should make general observation and estimations. Students will later find that the median for the period 4 is 91, and the median for period 5 is 82. Both class period test results are skewed to the right, toward the higher grades. Students will also notice there are 4 more students in period 5 compared to period 4.

2. If you were to make a histogram, what would be an appropriate window? In other words, what is the range of the data?

Answer: Answers can vary for the appropriate window, but a recommended viewing window would be 40 to 110. Period 4's grades range from 55 to 100. The scores in Period 5 range from 44 to 100.

3. How wide would you make the bins in the histogram? For example, if you make the bin width 10, how many period 4 and period 5 students are there with test scores of 60 to 69?

Answer: Answers will vary for the width of the bins. Students will be encouraged to explore various bin widths which could include 2, 3, 5, 10, etc. Period 4 has 3 scores in the 60s and period 5 has 2.

Period 4	Period 5
95	82
98	79
97	93
65	45
55	44
92	91
86	78
95	80
91	99
94	100
93	76
78	66
85	72
100	88
69	74
66	80
91	95
78	83
100	84
	96
	100
	84
	62

Tech Tip: Before beginning the activity, the list P4 and P5.8xl needs to be transferred to the students' calculators via handheld-to-handheld transfer or transferred from the computer to the calculator via TI-Connect.

Comparing distributions using histograms

Students will do the following in order to graph each data set using a histogram:

Press $\boxed{Y=}$ and clear or turn off any functions that are graphed. Press $\boxed{2nd} \boxed{Y=}$ for [STAT PLOT][STAT PLOT] to open the Stat Plot menu and choose Plot1.

Turn the plot **On**, choose the histogram icon for the plot type as shown on the right. Use $\boxed{ALPHA} \boxed{P4}$ to enter **P4** for the Xlist, or find the list in $\boxed{2nd} \boxed{[LIST]}$.

Press \boxed{WINDOW} to explore the data in the window you chose in question #2, or set it up with the window settings on the right. The **Xscl** determines the bin width. Press \boxed{GRAPH} to see the histogram.

Press \boxed{TRACE} and the arrow keys to see the contents of each bar.

Press \boxed{WINDOW} and change the **Xscl** value to adjust the width of the bars of the histogram. Experiment with different **Xscl** values.

Repeat these steps to graph **P5** as a histogram in Plot2.

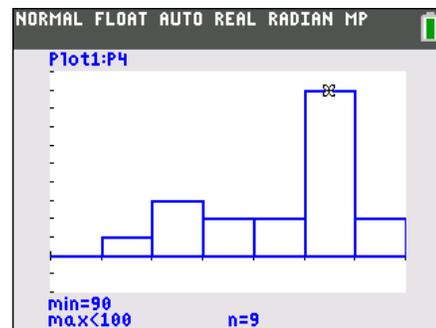
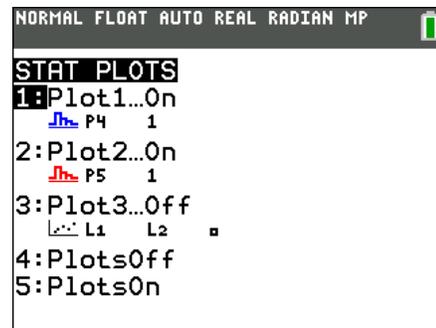
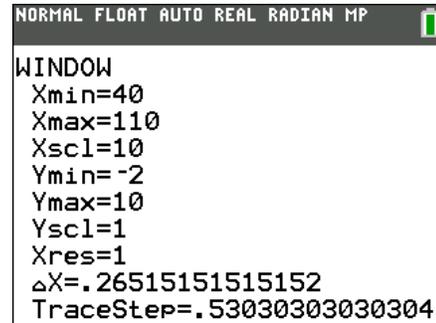
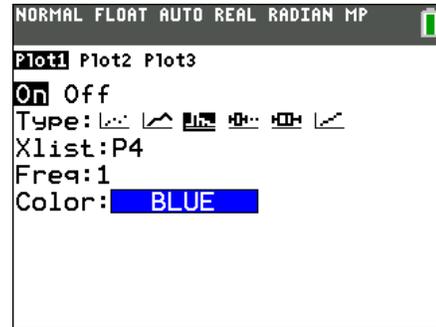
View the graphs together to compare them. Turn both Plot1 and Plot2 **On**.

- Why is it important for the scales to be the same when comparing the data?

Answer: It would be difficult to compare two histograms with different bin widths because a difference in the frequency of a bin could be the result of the bin width, not the actual data. Also it is difficult to determine how the maximum and minimums compare if the graph is in a different viewing window. To compare the test scores of the two periods, the scale of the window should be the same or it will be difficult to observe the differences.

- Compare the two histograms. Explain the shape and spread.

Answer: Both classes have scores that are skewed to the right. This means the classes have a greater frequency of students with high grades on the test. Period 5 scores are more spread out. The range of the period 5 data is larger than period 4.



6. Estimate the mean and median for each class. In which bin do you think the mean will be? Do you think the median will be to the left or right of the mean? Why?

Answer: The mean test score for both classes may be approximated to be in the bin that is greater than or equal to 80 and less than 90. The median will be to the right of, or greater than, the mean. The actual mean for period 4 is 85.7 and the median is 91. The mean for period 5 is 80.5 and the median is 82.

Students are given step-by-step instructions to create histograms. They are to first create the histogram for **P4**, the test scores for period 4, and then compare this to the histogram for the period 5, **P5**.

The **ZoomStat** command will graph the histograms on the same scale with the same bar width. However, when viewed separately, they will have different bar widths and scales. Also a limitation of **ZoomStat** is the default bin width. That is why it is recommended for the students to create their own window with an appropriate **Xscl**.

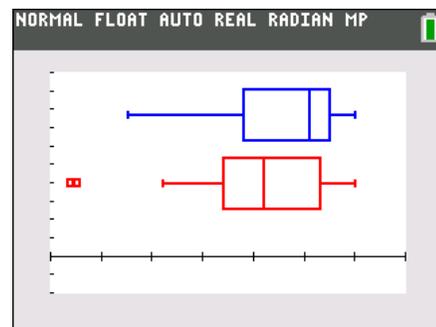
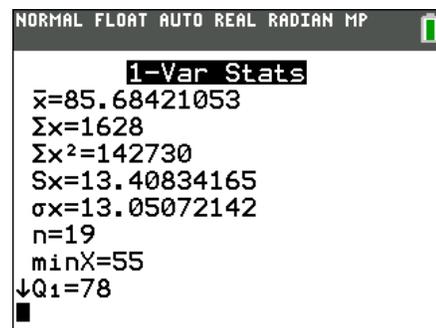
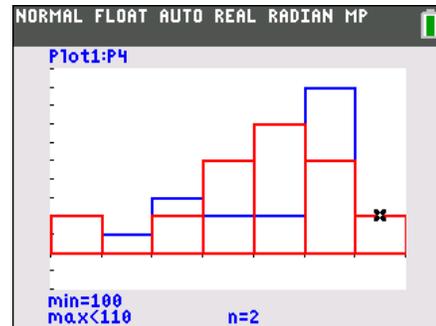
In their comparison, students should include discussion of the mean and median. First have them estimate each and then they can use the **mean(** and **median(** or one variable statistics commands to calculate them.

Remind students that since they are comparing real world data, their comparison should be in the context of the data.

Comparing distributions using box plots

Students create box plots using a similar set of steps as to those used when creating the histogram. After viewing the box plots separately, students can graph them together. The **ZoomStat** command will graph the box plots on the same scale. However, when viewed separately, they will have different scales.

Students are to include the five-number summary in their comparison.



Tech Tip: The five-number summary (minimum, lower quartile, median, upper quartile, and maximum) can be determined by pressing **TRACE**.

7. Record the actual mean and median for both classes and reflect on how your estimation compares.

Answer: The actual mean for period 4 is 85.7 and the median is 91. The mean for period 5 is 80.5 and the median is 82. Since the mean is less than the median, this is an indication that the test scores are skewed right. Students should also include a comparison with their estimation.

8. What would the graphs look like if the width of the bars is changed? Explore changing the bin width by changing the Xscl in **WINDOW**. You may also want to adjust the Ymax value. What effect does this have?

Answer: Answers will vary, but they should mention that if the width is decreased there will be fewer students who had those grades in each bin. If the bin width is increased there will likely be an increase in the frequency in the bins, i.e. more students will have scores on their test that are in that bin.

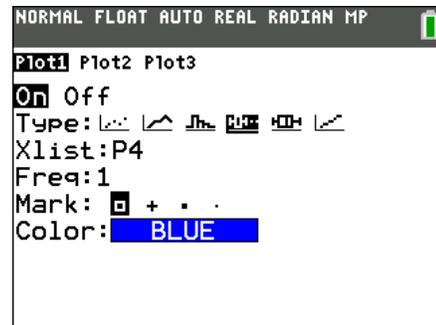
Comparing distributions using box plots

Sometimes box plots are more appropriate or useful to compare two data sets.

Graph each data set using a box plot.

Press **[2nd]** **[Y=]** to change the **[STAT PLOT]** for Plot1 and Plot2 to the first type of box plot as shown on the right. Use **ZoomStat** to view the graph in an appropriate window.

Press **[TRACE]** and the arrow keys to see the minimum, lower quartile, median, upper quartile, and maximum of each box plot.



Teaching Tip: Discuss questions that arise from the group work, including the concept of outliers in the dots in the 5th period box plot. Outliers are defined as points that lie more than one and half times the length of the box (the interquartile range) away from either end of the box. A good extension here would be to have students create two box plots of the period 5 test scores: one using the box plot that identifies outliers and one that does not.

9. What is the IQR, interquartile range, for period 4 and 5?

Answer: Period 4's IQR is 17, since $Q1 = 78$ and $Q3 = 95$. Period 5's IQR is 19, since $Q1 = 74$ and $Q3 = 93$.

10. What do the "dots" on period 5's test score box plot indicate?

Answer: The dots are the outliers. There are two students with grades less than 62.

11. Why are there outliers for period 5, but not for period 4?

Answer: For period 5, $1.5IQR$ is 28.5, and $Q1 - 1.5IQR = 45.5$, so any value smaller than that is an outlier. For period 4, $1.5IQR$ is 25.5, and $Q1 - 1.5IQR = 52.5$. There are no students with scores lower than 55 in period 4, so there are outliers the period 4, but not for in period 5.

12. What do you notice about the top 25% of both classes? Compare the two box plots, including shape, spread, and the five-number summary.

Answer: The bottom 25% for both classes is about the same. The shape and spread discussed above are still true; they are both skewed right.

The five-number summary for period 4 is $min=55$, $Q1=78$, $median=91$, $Q3=95$, and $max=100$.

The five-number summary for period 5 is $min=44$, $Q1=74$, $median=82$, $Q3=93$, and $max=100$.

13. What conclusions can you make about the test scores of the students in period 4 and 5? Each conclusion needs to be supported by a number or a graph.

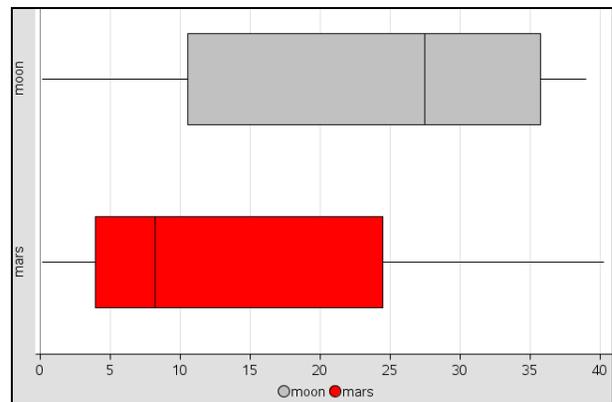
Answer: Answers will vary, but should summarize the earlier analysis.

Problem 2 – Moon and Mars Driving Distance

The graphs to the right come from data about the number of kilometers rovers have traveled on the Moon and Mars. The information comes from NASA JPL and was updated July 28, 2014.

14. Without exploring the actual data, what conclusions can you make from the box plot?

Answer: Answers should include that a Mars rover has traveled the furthest, but that most of the Mars vehicles did not travel as far as those on the Moon.



15. Find out how many rovers were manned and compare you conclusions with the details on the chart at <http://www.jpl.nasa.gov/news/news.php?release=2014-245>.

Answer: Students will see that 3 of the Lunar Rovers were driven by Americans. They should confirm their conclusions and discover that there were actually only 4 Mars rovers. This can be an interesting point of discussion. How do you get a five-point summary with only four data points?