Math Objectives

- Students will represent real-world data using a histogram and box and whisker diagram.
- Students will calculate measures of central tendency and spread.
- Students will compare distributions from two data sets and draw conclusions.
- Students will identify and analyze outliers in a data set.
- Students will try to make a connection with how to understand these topics in IB Mathematics courses and on their final assessments.

Vocabulary

- Histogram
- Skew
- Outlier
- Median
- Box and Whisker Diagram
- Interquartile Range

About the Lesson

- This lesson is aligning with the curriculum of IB Mathematics Applications and Interpretations and IB Mathematics Approaches and Analysis SL/HL
- This falls under the IB Mathematics Core Content Topic 4 Statistics and Probability:

  4.1: **(c)** Interpretation of outliers.

  4.2: **(b)** Histograms.

  (d) Production and understanding of box and whisker diagrams.

  4.3: **(a)** Measures of central tendency (mean, median, mode).

  (d) Measures of dispersion (interquartile range).

  As a result, students will:

- Apply this information to real world situations.

Teacher Preparation and Notes.

- This activity is done with the use of the TI-84 family as an aid to the problems.

Activity Materials

- Compatible TI Technologies: TI-84 Plus*, TI-84 Plus Silver Edition*, TI-84 Plus C Silver Edition, TI-84 Plus CE

  * with the latest operating system (2.55MP) featuring MathPrint^{TM} functionality.
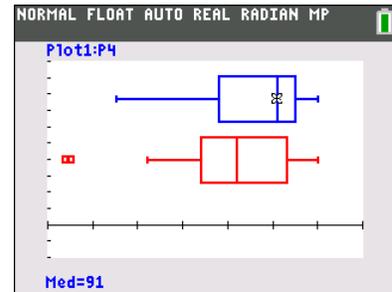
Tech Tips:

- This activity includes screen captures taken from the TI-84 Plus CE. It is also appropriate for use with the rest of the TI-84 Plus family. Slight variations to these directions may be required if using other calculator models.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

Lesson Files:
Student Activity
84CE-BestClass-Student.pdf
84CE-BestClass-Student.doc

In this activity, students create and explore a box and whisker diagram 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.
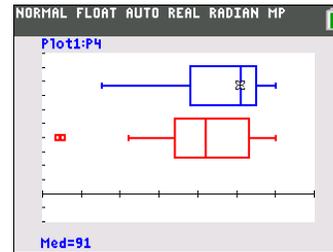
Problem 1 – Period 4 vs. Period 5 Test Results
Comparing distributions without graphing.

To the right are results from a recent test. The teacher wants to graphically display this data so the students can get a quick sense 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 and whisker diagrams.

Enter the test scores for period 4 and 5 data on the **STAT, EDIT** screen.

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.

   Solution: Students' answers should not be precise at this point. Both class period test results are skewed right toward higher grades. Students should notice that there are 4 more students in period 5 compared to period 4.

2. If you were to make a histogram, describe an appropriate window. In other words, find the range of the data.

   Solution: Answers vary. Recommended window would be 40 to 110. Period 4's range from 55 to 100 and period 5's from 44 to 100.

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

   Solution: Answers vary. Students should explore bin widths to include 2, 3, 5, 10, etc. Period 4 has 3 scores and period 5 has 2 scores in the 60s.

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

Comparing distributions using histograms

Graph each data set using a histogram.

Press **y=** and clear or turn off any functions that are graphed. Press **2ⁿᵈ, y=** for **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 **ALPHA** to enter **P4** for the Xlist, or find the list in **2ⁿᵈ, [LIST]**.

Press **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 **GRAPH** to see the histogram.

Press **TRACE** and the arrow keys to see the contents of each bar.

Press **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**.

4. Explain why it is important for the scales to be the same when comparing the data.

    Solution: 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.
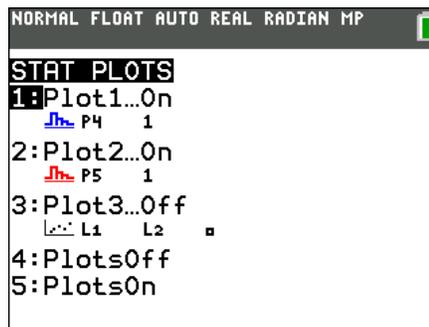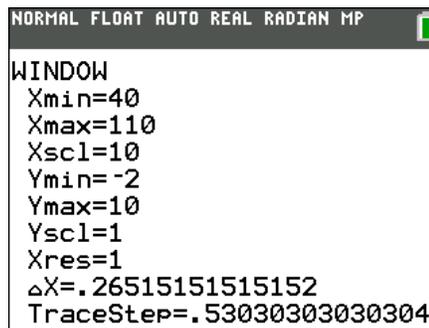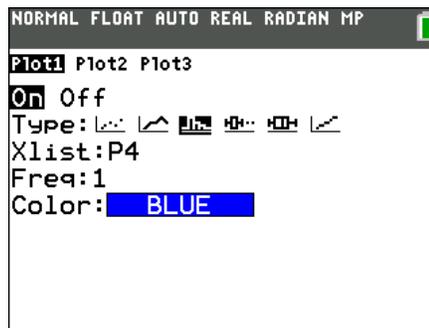
5. Compare the two histograms. Explain the shape and spread.

   Solution: 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 team. Explain which bin you think the mean will be in. Discuss with a classmate if you think the median will be to the left or right of the mean. Explain.

   Solution: 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.

Now compare your estimate of the mean and median to the actual values. Press **STAT**, arrow to the right once to select **CALC** and press 1 to select one variable statistics. Press **2ⁿᵈ,[List]** to select P4 for the 4th period. Then arrow down to **Calculate** and press **enter**. The symbol for mean is $\bar{x}$. Arrow down to see the median, abbreviated **Med**. Repeat for P5, 5th period data.
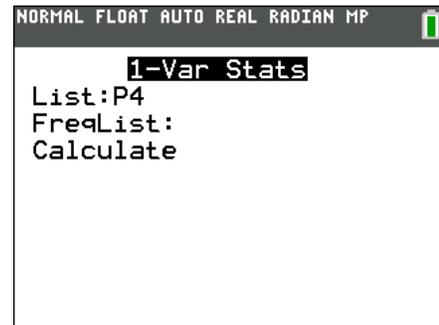
Students are given step-by-step instructions to create histograms. They are to first create the histogram for **P4**, the test scores for period 4, and then compare this to the histogram for the period 5, **P5**.

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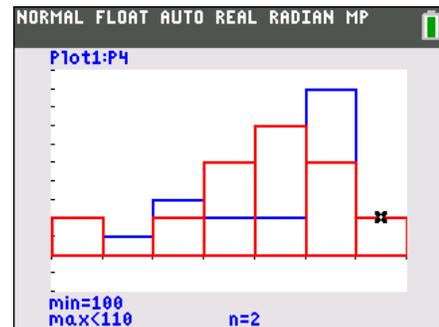
In their comparison, students should include discussion of the mean and median. First have them estimate each and then they can use the **mean(** and **median(** or one variable statistics commands to calculate them.

Remind students that since they are comparing real world data, their comparison should be in the context of the data.
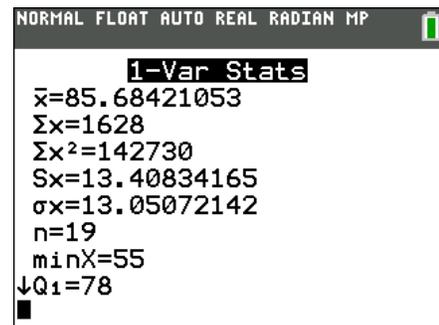
7. Record the actual mean and median for both teams and reflect on how your estimation compares.

   Solution: 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. Describe what the graphs would look like if the width of the bars is changed. Explore changing the bin width by changing the **Xscl** in **WINDOW**. Explain what effect this would have.

   Solution: 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 and whisker diagrams

Sometimes, a box and whisker diagram is more appropriate or useful to compare two data sets.

Graph each data set using a box and whisker diagram.

Press **2nd, y=** to change the **[STAT PLOT]** for Plot1 and Plot2 to the first type of box and whisker diagram 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 and whisker diagrams.
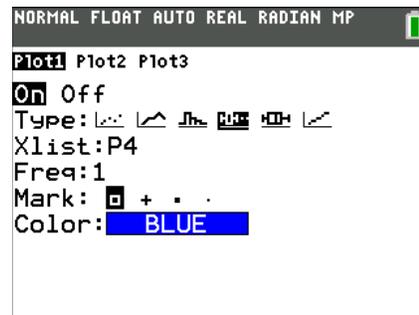
Teacher Tip: Discuss questions that arise from the group work, including the concept of outliers in the dots in the 5th period box and whisker diagram. Outliers are defined as points that lie more than one and half times the length of the box (the interquartile range) away from either end of the box. A good extension here would be to have students create two box and whisker diagrams of the period 5 test scores: one using the box and whisker diagram that identifies outliers and one that does not.

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

Students are to include the five-number summary in their comparison.
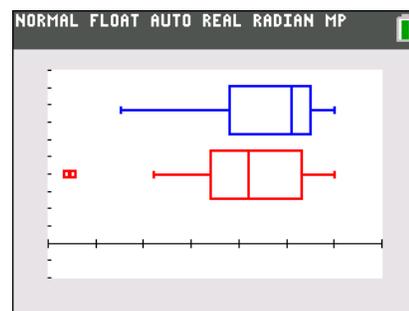
Tech Tip: The five-number summary (minimum, lower quartile, median, upper quartile, and maximum) can be determined by pressing **Trace**.

9. Find the IQR, interquartile range, for period 4 and 5.

    Solution: 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. Explain what the "dots" on period 5's test score box and whisker diagram indicate.

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

11. Explain why there are outliers for period 5, but not for period 4.

    Solution: 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. Explain what you notice about the top 25% of both classes. Compare the two box and whisker diagrams, including shape, spread, and the five-number summary.

    Solution: 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. Explain the conclusions you can 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.

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

Further IB Application

Kevin is trying to improve his putting. For the forty rounds of golf he played last year, the number of one putt holes Kevin has in a given 18 hole round of golf was recorded. The data was organized in a box and whisker diagram as shown below.
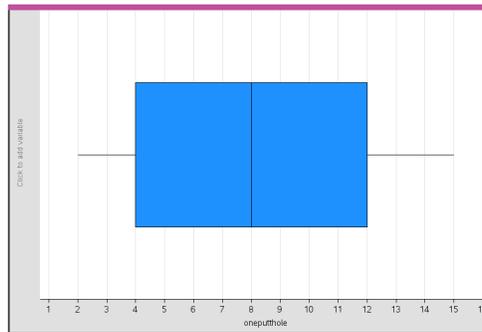
For this data, write down

(a) (i) The minimum number of one putt holes for a round out of the 40 rounds Kevin played last year

Solution: 2 one putt holes

(ii) The lower quartile

Solution: 4 one putt holes

(iii) The median.

Solution: 8 one putt holes

(b) Kevin claims that this box and whisker diagram can be used to infer that the percentage of rounds with 5 one putt holes is smaller than the percentage of rounds with 9 one putt holes. State whether Kevin is correct. Justify your answer.

Solution: Each of the sections of the box and whisker diagram represent 25% of the data, or the diagram does not show individual percentiles only the quartiles and since no raw data is given, Kevin is incorrect and no such inference can be made.

Teacher Tip: Throughout this activity, the students are asked to discuss with classmates and explain how they achieved their answers. This is a wonderful opportunity to create a student led classroom. As you float around the room, listen to what they are saying, add to their discussions, and give them leading questions to see how they respond.

Teacher Tip: Basic understanding of the TI-84 CE is needed to navigate through the screens and commands with respect to statistics. It may be a good idea to review these before the activity.

**Note: This activity has been developed independently by Texas Instruments and aligned with the IB Mathematics curriculum, but is not endorsed by IB™. IB is a registered trademark owned by the International Baccalaureate Organization.*