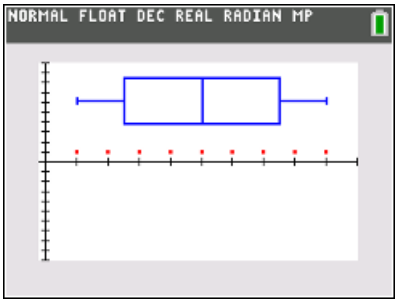
Give Me Five!
Student Activity

Name _____
Class _____

In this activity, students will discuss and describe the center and spread of a univariate data set by way of a five number summary and visually by a box & whisker diagram. Students will then apply this knowledge to real life applications to enhance their ability to understand this math in statistical data analysis.



Introduction

A univariate set of data is a list of numbers that describes the different value of a variable characteristic across a range of different units. For example, if a study involved finding out the height of a range of people, each person whose height is measured is statistically considered to be a 'unit'. Height is the characteristic that varies (variable) and the list of height measurements is called the data.

When describing a group of data, there are generally two main types of things to consider:

a) Measure of center – this is a single value that could be used as a representative of the entire data set (e.g. mean, median, mode)

b) Measure of spread – this is a number that indicates how spread out the data are (e.g. standard deviation, range, inter-quartile range)

Problem 1 – The Fantastic Five

A five number summary is a convenient way of describing a set of data as it provides us with information about both center and spread. Consider the set of data: {1, 2, 3, 4, 5, 6, 7, 8, 9}. We can see that the numbers are already ordered from lowest to highest.

1. Find the *minimum* value in the data set. We call this value **MinX**.

2. Find the *maximum* value in the data set. We call this value **MaxX**.

3. Find the *middle* value in the data set. We call this value **MedianX**.

4. Look at the numbers that are less than the Median, find the median of *this* set of numbers. Discuss with a classmate what the median would be if this data set was an even number of data and if this data set was an odd number of data. Find the name you would give this piece of data.

5. Look at the numbers that are less than the Median, find the median of *this* set of numbers. Discuss with a classmate what the median would be if this data set was an even number of data and if this data set was an odd number of data. Find the name you would give this piece of data.

6. You have found Q_1 and Q_3. Discuss with a classmate what Q_2 is.

The five numbers that are your answers to questions 1 to 5 are called the **five number summary**. Usually the five number summary is written in the order MinX, Q_1, Median, Q_3, MaxX. The median (Answer to question 3) is the measure of center. The other numbers provide indications of spread.

- MaxX minus MinX is the **Range**.
- Q_3 minus Q_1 is the **Inter-Quartile Range (IQR)**.

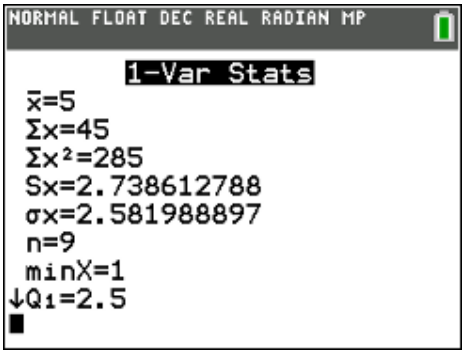
7. Find the *range* for the data set {1, 2, 3, 4, 5, 6, 7, 8, 9}.

8. Find the *Inter-Quartile Range* for the data set {1, 2, 3, 4, 5, 6, 7, 8, 9}.

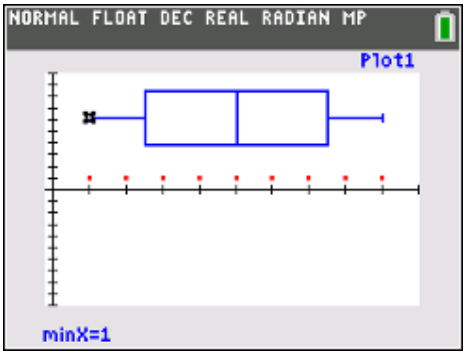
Problem 2 – Automatic Calculation of the Five Number Summary

On the handheld, press **stat**, **1: edit** and enter the data set {1, 2, 3, 4, 5, 6, 7, 8, 9} into L_1 and the data set {1,1,1,1,1,1,1,1,1} into L_2.

The data values are now entered into a List under the column title of L_1. Calculate the summary statistics for the data by pressing **stat**, **CALC**, **1: 1-VAR STATS** (list is L_1 and press calculate) to display the statistics we will be referencing in this activity.



Next, create a graphical representation of the data set, called a Box Plot or a Box & Whisker Diagram. Press **2nd y =** (**statplot**). Under Plot1, turn it on, select the first box plot, and make sure your xlist is L_1. Under Plot2, turn it on, select the scatter plot and make sure your xlist is again L_1 and your ylist is L_2 to create a dot plot beneath the Box Plot.



If you press trace and the left and right arrows, the five numbers of the five number summary will be revealed. Note that they are in line with corresponding numbers on the scale below it.

Go back to your data list (stat, edit) and change the final value from a nine to a ten and return to your Box Plot.

1. Explain why Q_1, the median and Q_3 do not change when the data point (9) is increased.

2. Keep changing this final value. Explain what happens to the whisker when this data point is moved further and further away from the rest of the data. Find at what value, approximately, this significant change occurs.

3. Return to your lists page (**stat**, **edit**) and enter the data set {9, 3, 8, 5, 7, 4, 1, 6, 2} into L_3. Discuss with and state the affects this may have on the Box Plot and statistics calculations. Explain why you think this is so.

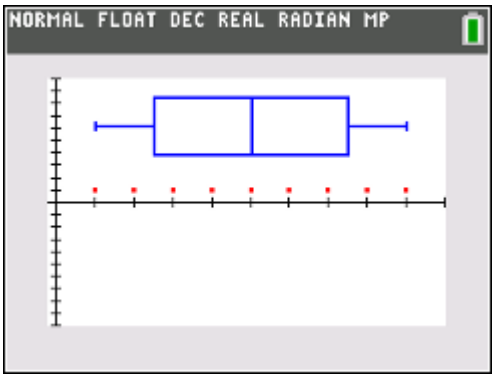
4. Compute the five number summary of the data set, as you did at the start of problem 2, by placing this data on your lists page under L_4: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. Validate the five number summary by hand.

5. Use your answer to the previous question to find a data set that has a five number summary made up entirely of integers.

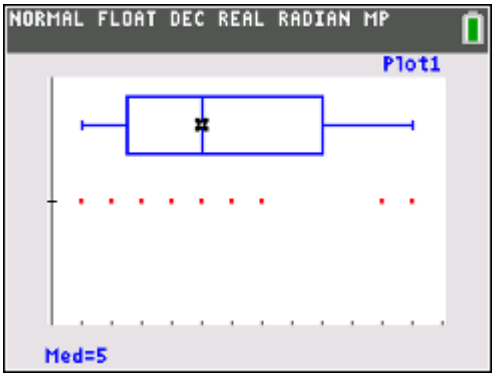
Problem 3 – Consideration of shape and skew

So far the data sets we have considered have been **symmetrical**. That is, the Box Plot is geometrically symmetrical and has a vertical line of symmetry at the median. This means that Q_1 is as far below the median as Q_3 is above it and MinX is as far below the median as MaxX is above it. You may also have noticed that the **mean** value (the first value shown when finding the 1-VAR STATS) is always the same as the median.

Using the original data set: {1, 2, 3, 4, 5, 6, 7, 8, 9}, note that the median value is 5, as also is the mean. Looking at the Box Plot for these data, notice that it is perfectly symmetrical.



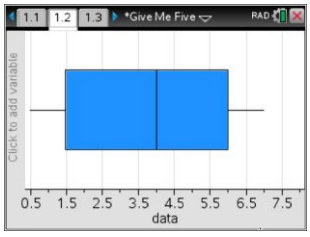
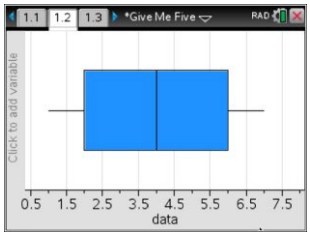
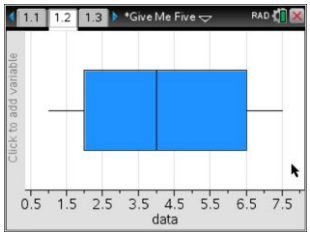
Going back to your lists page, replace the final two values of L_1 (8, 9) with 11 and 12. Return to the graph. Q_3 will move up to about 9. Notice that the distribution is now no longer symmetrical. The part of the box that is between the median and Q_3 is bigger than the part between the median and Q_1. The distribution is now said to be **positively skewed** or **skewed right**.



Notice also that, although the median is still 5, the mean value has moved up to about 5.6. Restore

the original data set and repeat this process to show a **negative skew** or **skew left**.

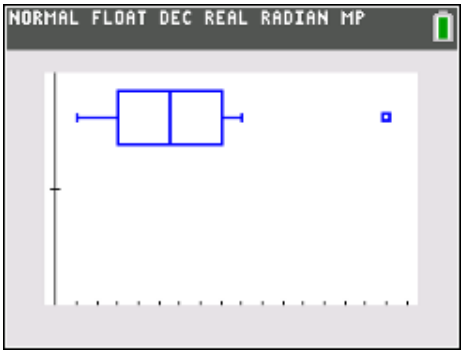
1. Match each of the following Box Plots with its matching description of symmetry and comment about measure of center.

Box Plot	Description of Shape	Comment on measure of center
	Symmetrical	Mean > Median
	Positively skewed	Mean < Median
	Negatively skewed	Mean = Median

Outliers and Fences

Restore your data list to the set: {1, 2, 3, 4, 5, 6, 7, 8, 9}.

Now add in a 10th value to the list. Make this value 16. Observe the corresponding Box Plot. Notice that the value 16 doesn't appear within the main box or whisker, but is shown as a dot on its own. This is because the score 16 is so far away from the other data points that it is considered to be an **outlier**.



Experiment by replacing the 16 with values that are closer to the original data set. Try replacing it with 14, 13, 12, 11, 10.

A numerical value that determines the threshold for outliers can be computed and is referred to as the **upper fence value** where the outlier is above the median and **lower fence value** where the outlier is below the median. The upper and lower fences are defined by using the Inter-Quartile Range (IQR).

Upper Fence Value = Q3 + 1.5 x IQR

Lower Fence Value = Q1 - 1.5 x IQR

Example

If you have a set of 8 scores {1, 2, 3, 4, 5, 6, 7, 8}, such that $Q_1 = 2.5$, $Q_3 = 6.5$ and the IQR = 4.

Upper fence = 6.5 + 1.5 x 4
= 12.5

Lower Fence = 2.5 – 1.5 x 4
= -3.5

2. Find the Upper and Lower Fences for the data set {1, 2, 3, 4, 5, 6, 7, 8, 9}. Find and explain what data values would be considered outliers if added to this data set.

3. Discuss with a classmate and explain how it is possible to calculate the IQR whilst a single outlier is changed.

Further IB Application

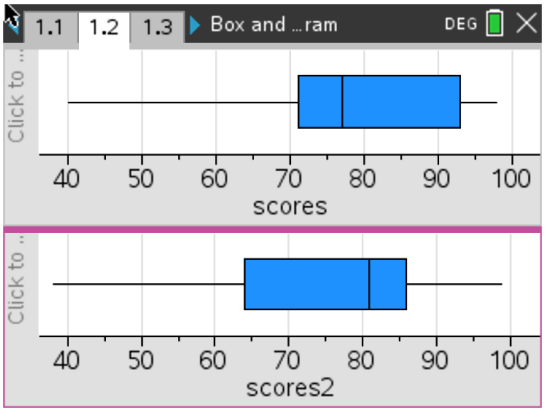
The scores of a mathematics test given to period 1 are shown below.

40, 62, 65, 71, 73, 74, 75, 77, 80, 90, 92, 93, 96, 97, 98

For the data, the lower quartile is 71 and the upper quartile is 93.

(a) Show that the test score of 40 would not be considered an outlier.

The same mathematics test was given to period 2 and the box and whisker diagram showing their scores (**scores2**) and comparing them to the scores of period 1 (**scores**) are below.



A fellow mathematics teacher looks at the box and whisker diagrams and believes that period 2 performed better than period 1.

(b) Using the diagrams above, state one reason that may support the mathematics teacher's opinion and one reason that may counter it.