Five Figure Summary

Aim

- To describe the centre and spread of a univariate data set by way of a 5-figure summary and visually by a box & whisker plot.

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 centre – 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)

The Fantastic Five

A 5-figure summary is a convenient way of describing a set of data as it provides us with information about both centre 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.

Question: 1

What is the *minimum* value in the data set? We call this value **MinX**.

Question: 2

What is the *maximum* value in the data set? We call this value **MaxX**.

Question: 3

What is the *middle* value in the data set? We call this value **MedianX**.

Question: 4

Look at the numbers that are less than the Median[1], what is the median of *this* set of numbers?

Hint: As there are only four numbers in this group, you have to work out what number is exactly half way between the middle two digits, for this data set it will be a fraction, we call this value Q_1 or the **Lower Quartile**.

---

[1] Calculating Q_1 (or Q_3): There are two options for these calculations. Method 1: {**1, 2, 3, 4**, 5, 6, 7, 8, 9} → Q_1 = 2.5 … or …
Method 2 {**1, 2, 3, 4**, **5**, 6, 7, 8, 9} → Q_1 = 3. TI-Nspire uses the first calculation method where the median is not included.

Author: Brian Lannen

TEXAS INSTRUMENTS

Question: 5

Look at the numbers that are more than the Median, what is the median of *this* set of numbers?

Hint: As there are only four numbers in this group, you have to work out what number is exactly half way between the middle two digits, for this data set it will be a fraction, we call this value **Q3** or the **Upper Quartile**.

Question: 6

We have found Q_1 and Q_3. What happened to Q_2?

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

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

Question: 7

What is the *range* for the data set {1, 2, 3, 4, 5, 6, 7, 8, 9}?

Question: 8

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

Automatic Calculation of the Five Figure Summary

Open the TI-Nspire file:

Give Me Five.tns.

Observe the data set {1, 2, 3, 4, 5, 6, 7, 8, 9} on page 1.2.

The data values are entered into a Lists & Spreadsheet page under the column title of 'data'. Summary statistics for the data have been calculated and displayed in the next two columns.


Page 1.3 provides a graphical representation of the data set. This type of graph is called a Box Plot or a Box & Whisker Plot. A dot plot is included beneath the Box Plot.

If you scroll your calculator cursor over this graph, the five figures of the 5-figure summary will be revealed. Note that they are in line with corresponding numbers on the scale below.
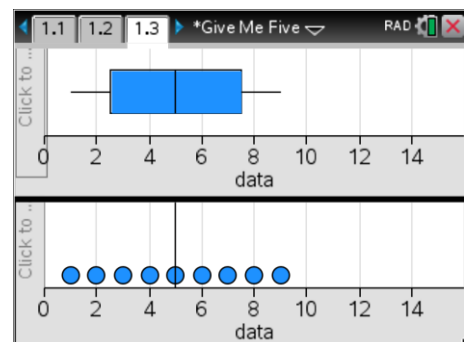
You can navigate between the two Data and Statistics applications by using **Ctrl** + **Tab**, similar to ALT + Tab on a Windows™ platform.

Author: Brian Lannen

TEXAS INSTRUMENTS

Move the cursor over the top of the data point with the largest value in the dot plot. (9) Grab the point (Click and hold) and move it to the right.
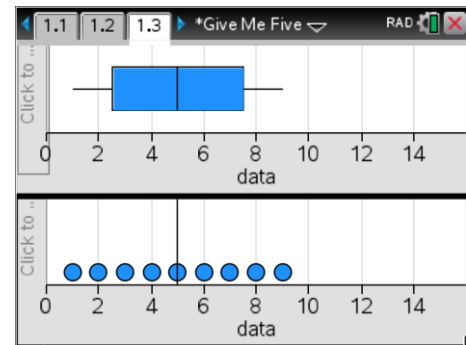
Question: 9

> Explain why Q_1, the median and Q_3 do not change when the data point (9) is increased.
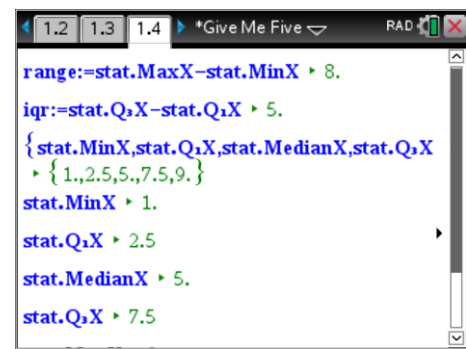
Question: 10

> What happens to the whisker when this data point is moved further and further far away from the rest of the data?   At approximately what value does this significant change occur?

Navigate to page 1.4. This is Dynamic Notes page that is set up to show the **range** and the **interquartile range** of the data set from page 1.2.

It shows the 5-Figure summary as a set of 5 values in brackets {}.

The five values are also shown on separate lines below this.

The best thing about this automatic calculator is that if you go back to page 1.2 and make changes to the data set, the Box Plot and the Dynamic Notes 5-Figure Summary will automatically change to reflect the new data set.

Question: 11

> Navigate to page 1.2 and replace the data set {1, 2, 3, 4, 5, 6, 7, 8, 9} with {9, 3, 8, 5, 7, 4, 1, 6, 2}. What affect does this have on the Box Plot and statistics calculations? Why do you think this is so?

Question: 12

> Use the automatic calculator (Page 1.4) to compute the 5 Figure Summary of the data set: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. Validate the 5 Figure Summary by hand.

Question: 13

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

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 (indicated in column C of 1.2) is always the same as the median.

Author: Brian Lannen

TEXAS INSTRUMENTS

If necessary, adjust the data set on Page 1.2 so that it is equal to: {1, 2, 3, 4, 5, 6, 7, 8, 9}. Note that the median value is 5, as also is the mean. Now move to Page 1.3 and observe the Box Plot for these data, it is perfectly symmetrical. Click on the MaxX value at the end of the right-hand whisker. The MaxX value (9) will be highlighted and also the value (8).
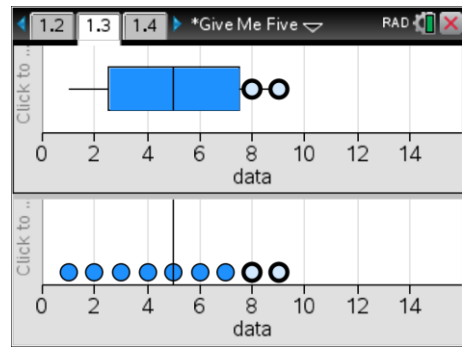
Drag the whisker to the right so that MaxX is at about 12.

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**.
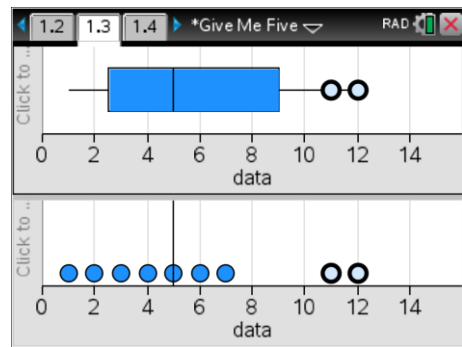
Notice also that, although the median is still 5, the mean value has moved up to about 5.6. Restore to the original data set and repeat this process to show **negative skew**.
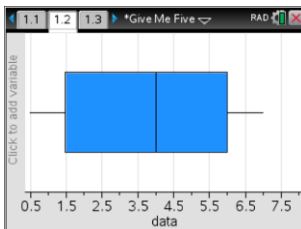

Question: 14

Match each of the following Box Plots with its matching description of symmetry and comment about measure of centre.
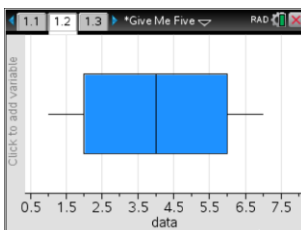
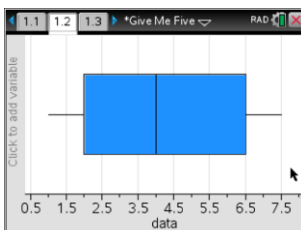| Box Plot	Description of Shape	Comment on measure of centre
	Symmetrical	Mean > Median
	Positively skewed	Mean < Median
	Negatively skewed	Mean = Median

Author: Brian Lannen

TEXAS INSTRUMENTS

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 15 with values that are closer to the original data set. Try replacing it with 14, 13, 12, 11, 10.
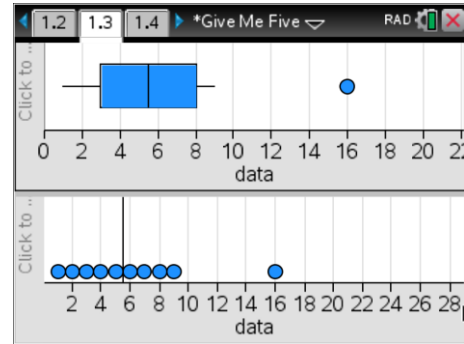
A numerical value that determines the threshold for outliers can be computed and is referred to as the **upper fence value** where the outlier is above the median and **lower fence value** where the outlier is below the median. The upper and lower fences are defined by using the Inter-Quartile Range (IQR).

Upper Fence Value = Q3 + 1.5 x IQR

Lower Fence Value = Q1 - 1.5 x IQR

Example

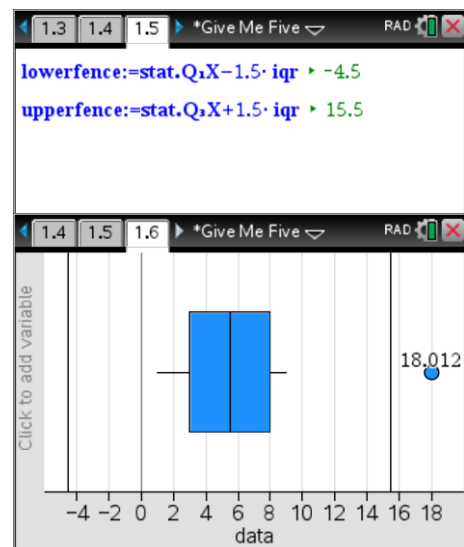If you have a set of 8 scores {1, 2, 3, 4, 5, 6, 7, 8}, such that $Q_1 = 2.5$, $Q_3 = 6.5$ and IQR = 4.

Upper fence = 6.5 + 1.5 x 4

= 12.5

Lower fence = 2.5 – 1.5 x 4

= -3.5

Page 1.5 is a Dynamic Notes page that shows the calculation of Lower and Upper Fences.

Page 1.6 Shows the Box Plot again, this time with the fences displayed as vertical lines.

Drag the outlier to the fence and observe what happens.

Question: 15

Explain how it is possible to calculate the IQR whilst a single outlier is changed.

Author: Brian Lannen

TEXAS INSTRUMENTS