# ESSENTIAL Mathematics 2

## Week 5 NOTES and exercises

**Back-to-Back Stem and Leaf Plots**

We now know how to construct a stem-and-leaf plot for a set of data. We can extend a stem-and-leaf plot so that it shows two sets of data. This is useful when one we wish to compare one set of data against another.

**Example 1**

The girls and boys in Grade 4 at Kingston Primary School submitted projects on the Olympic Games. The marks they obtained out of 20 are given below.



Display the data on a back-to-back stem plot.

1. Identify the highest and lowest scores in order to decide on the stems.

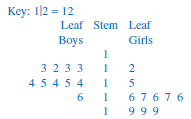
Use a stem of 1, divide into fifths.

Highest score = 19

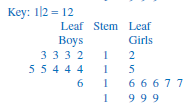
Lowest score = 12

ie 0-1, 2-3, 4-5, 6-7, 8-9

2. Create an unordered stem plot first. Put the boys’ scores on the left, and the girls’ scores on the right.



3. Now order the stem plot. The scores on the left should increase in value from right to left, while the scores on the right should increase in value from left to right.



The back-to-back stem plot allows us to make some visual comparisons of the two distributions. The centre of the distribution for the girls is higher than the centre of the distribution for the boys. The spread of each of the distributions seems to be about the same. For the boys, the marks are grouped around the 12–15 marks; for the girls, they are grouped around the 16–19 marks. On the whole, we can conclude that the girls obtained better marks than the boys did.

To get a more precise picture of the centre and spread of each of the distributions we can use the summary statistics we have used before. Specifically, we are interested in:

1. the mean and the median (to measure the centre of the distributions), and

2. the interquartile range (to measure the spread of the distributions).

Boys Girls

Median = 14 Median = 16.5

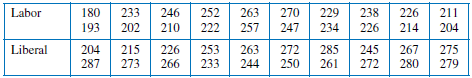
Mean = 13.9 Mean = 16.6

Q3 – Q1 = 2 Q3 – Q1 = 3

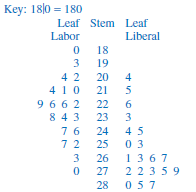
Remember: Q3 – Q1 is the interquartile range

**Example 2**

The number of ‘how to vote’ cards handed out by various Australian Labor Party and Liberal party volunteers during the course of a polling day is shown below.



The stem and leaf plot becomes



For the Labor volunteers: For the Liberal volunteers:

Mean = 227.9 Mean = 257.5

Median = 227.5 Median = 264.5

Interquartile range = 36 Interquartile range = 29.5

From the stem plot we see that the Labor distribution is symmetric and therefore the mean and the median

are very close, whereas the Liberal distribution is skewed towards to high 200’s.

Since the distribution is skewed, the median is a better indicator of the centre of the distribution than

is the mean.

Comparing the medians therefore, we have the median number of cards handed out for Labor at 228

and for Liberal at 265, which is a big difference.

The data, as shown by the interquartile range, is slightly more spread for Labor as compared to the Liberals

In essence, the Liberal party volunteers handed out a lot more ‘how to vote’ cards than the Labor party

volunteers did.

**Exercise Set 1**

Q1. The marks (out of 50), obtained for the end-of-term test by the students in German and French classes are given below. Display the data on a back-to-back stem plot.



Q2. The birth masses of 10 boys and 10 girls (in kilograms, to the nearest 100 grams) are recorded in the table below. Display the data on a back-to-back stem plot.



Use the median and mean to make a comparison between the weights of the girls and boys.

Girls Boys

Median = Median =

Mean = Mean =

Statement:

Q3. The number of delivery trucks making deliveries to a supermarket each day over a 2-week period was recorded for two neighbouring supermarkets —supermarket A and supermarket B. The data are shown below. Construct a back-to-back stem and leaf plot.



Use the median and mean to make a comparison between the number of trucks delivering to A and B.

A B

Median = Median =

Mean = Mean =

Statement:

Q4. These are the results of two Year 11 classes in their final mathematics exam.

Group A



Group B



a) Draw a back-to-back stem and leaf plot for this data.

b) Find the median score for each class.

c) Find the range for each class.

d) Are there any outliers in either class? If so, state the outliers.

e) If one class is the top class and one class is the middle class, which is which? Why?

Q5. Mr Farley, the school Principal, is concerned about the absences in Year 11 and Year 12. The following shows the number of absentees per day in Year 11 and Year 12 over a 4 week period.



a) Draw a back-to-back stem and leaf plot for this data.

b) Find the median score for each year.

c) Find the range for each year.

d) Are there any outliers in either class? If so, state the outliers.

e) State one similarity between the two scores.

f) State one difference between the two scores.

g) Mr Farley believes there is more absenteeism in Year 11 than in Year 12. What does this information indicate? Is more information needed?

**Boxplots**

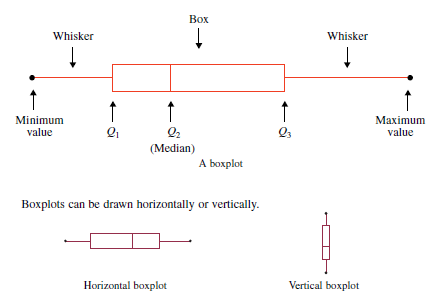
The five number summary statistics are as follows;

Lowest score, lower quartile Q1, median Q2, upper quartile Q3, highest score.

This information can be illustrated very neatly in a special diagram known as a *boxplot* (or *box-and-whisker* diagram). The diagram is made up of a box with straight lines (whiskers) extending from opposite

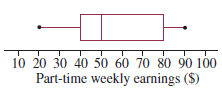
sides of the box.

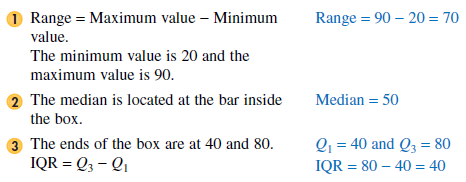
A boxplot displays the minimum and maximum values of the data together with the quartiles and is drawn with a scale. The length of the box gives us the interquartile range. A boxplot gives us a very clear visual display of how the data are spread out.



**Example**

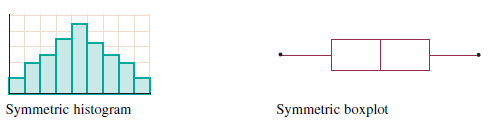
The boxplot below shows the distribution of the part-time weekly earnings of a group of Year-11 students. Write down the range, the median and the interquartile range for these data.





**Skewness**

In the figure below a symmetric distribution is represented in the histogram and in the boxplot. The characteristics of this boxplot are that the whiskers are about the same length and the median is located about halfway along the box.



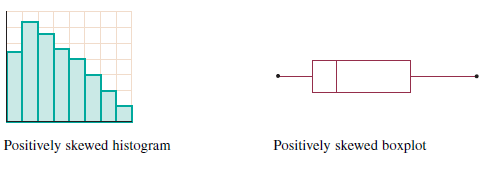
The figure below shows a negatively skewed distribution. In such a distribution, the data peak to the right on the histogram and trail off to the left. In corresponding fashion on the boxplot, the bunching of the data to the right means that the left-hand whisker is longer and the right-hand whisker is shorter; that is, the lower 25% of data are sparse and spread out whereas the top 25% of data are bunched up.

The median occurs further towards the right end of the box.



In the figure below we have a positively skewed distribution. In such a distribution, the data peak to the left on the histogram and trail off to the right. In corresponding fashion on the boxplot, the bunching of the data that the left-hand whisker is shorter and the right-hand whisker is longer; that is, the upper 25% of data are sparse and spread out whereas the lower 25% of data are bunched up.

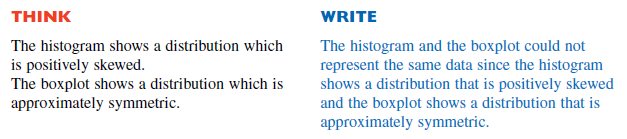
The median occurs further towards the left end of the box.



**Example**

Explain whether or not the histogram and the boxplot shown below could represent the same data.



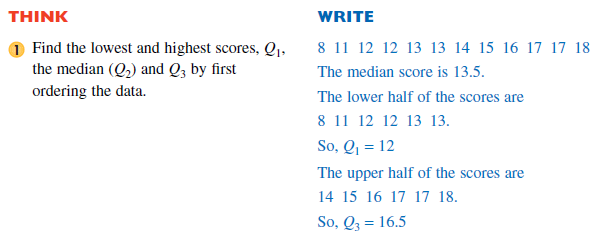


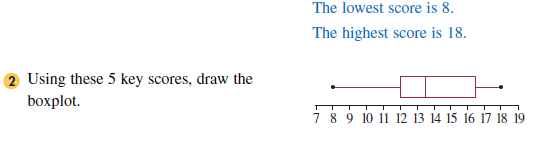
**Example**

The results out of 20 of oral tests in a Year-12 Indonesian class are:

15 12 17 8 13 18 14 16 17 13 11 12

Display these data using a boxplot.

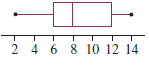




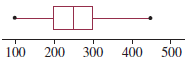
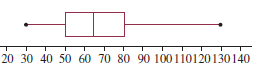
**Exercise Set 2**

Q1. For the boxplots shown, write down the range, the interquartile range and the median of the distributions which each one represents.

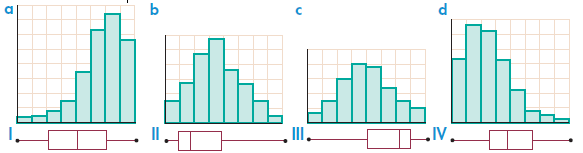
a) b)

c) d)

Q2. Each of the histograms shown below is labelled with a letter and each of the boxplots is labelled with a number. Match each histogram with a boxplot which could show the same distribution.



Q3. For each of the following sets of data construct a boxplot.

a) 3 5 6 8 8 9 12 14 17 18

b) 3 4 4 5 5 6 7 7 7 8 8 8 9 9 10 10 12

Q4. The maximum daily temperatures (in °C) for the month of October in Melbourne are:



Represent this data in a boxplot.

**Parallel Boxplots**

In statistics there are many opportunities to compare two sets of data. We can compare sets of data by drawing two or more boxplots using a common scale.

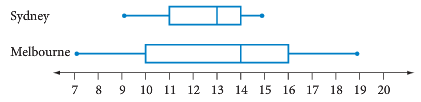
**Example**

The following two five number summaries for Sydney and Melbourne describe the number of rainy days per month over two years.

Sydney: 9, 11, 13, 14, 15

Melbourne: 7, 10, 14, 16, 19

The boxplots are placed on a common scale.



Median for Sydney = 13 Median for Melbourne = 14

Interquartile range for Sydney = 14 – 11 = 3 Interquartile range for Melbourne = 16 – 10 = 6

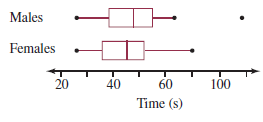
Comparison: Melbourne has more rainy days per month. Its median is higher and half its scores are above 14 compared to one quarter of the scores for Sydney.

Sydney has a more consistent pattern of rainy days because its range and interquartile range are smaller than Melbourne’s.

**Exercise Set 3**

Q1. A concentration test was carried out on 40 students in Year 12 across Australia. The test involved the use of a computer mouse and the ability to recognise multiple images. The less time required to complete the activity, the better the student’s ability to concentrate.

The data are shown by the parallel boxplots below.



a) Identify one similarity and one difference between the concentration spans of boys and girls.

b) Find the interquartile range for the boys and girls.

c) Comment on the existence of an outlier for the boys.

Q2. Rigby and Alex are in different Year 11 Math classes. The following five number summaries are for half-yearly exams in each class.

Rigby’s class: 48, 64, 75, 87, 96

Alex’s class: 47, 57, 69, 80, 96

a) Draw a double boxplot of these summaries.

b) What is the range for each class?

c) Both Alex and Rigby scored 85% in their half-yearly exams. Who has performed better in relation to their own class? Justify your answer.

d) Can we calculate the mean from the information given? Expain.

Q3. The Mayor of Middleton claims that his town is safer for drivers than the nearby town of Blakewell. To test this, James measured the speed (in km/hr) of a sample of 20 cars in each town.



a) Arrange each set of data in order.

b) Find the median for each town.

c) Find the mean for each town.

d) Find the mode for each town.

e) Complete the five number summary for each set of data.

f) Construct double boxplots for this data.

g) Is the Mayor of Middleton correct? Justify your answer.