Week 5
Term 3
2024

HAWKER COLLEGE
Engage | Inspire | Achieve

EM2
Representing and comparing data

Goals

Goals for this week:

- compare back-to-back stem plots for different data-sets
- complete a five number summary for different data-sets
- construct box plots using a five number summary
- compare the characteristics of the shape of histograms using symmetry, skewness and bimodality

Girls		Boys
7, 8, 2, 2, 1	1	5, 8
3, 3, 3, 2	2	2, 2, 3, 6
5, 4, 3	3	4, 5, 5, 5
7, 5, 4	4	0, 0, 2, 7, 9
1, 1, 0	5	0, 0, 1

Theoretical Components

STEP 1

Resources:
PDF file: Week 5 Notes and Exercises

This Week:
We will be learning:

- Comparing data using back to back stem and leaf plots
- Using five figure summaries to construct boxplots
- Comparing sets of data using parallel box plots

Practical Components

STEP 2
Read through Week 5 Notes and Exercises for instructions on what to do.

There are 3 Exercises in this booklet. Read any worked examples before you begin.

Remember to regularly check Google Classroom for messages.

Portfolio Task

STEP 3

Complete the task at the end of the brief and submit your weekly work. 😊

Other

If you haven't completed the work from last week, you need to complete it and hand it in.

ESSENTIAL MATHEMATICS 2

WEEK 5 – COMPARING DATA SETS

How are we ever going to use this?

- When comparing products for sale
- When examining performances in a variety of sports
- To interpret data presented in the media

It is important to be able to compare data sets because it helps us make conclusions or judgements about the data. Some sets of data are easily compared directly because they are measuring similar things, but sometimes we may need to compare data sets that are quite different. When comparing two or more sets of data, it may be helpful to use graphical displays of the data.

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:

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.

Girls' marks	16	17	19	15	12	16	17	19	19	16
Boys' marks	14	15	16	13	12	13	14	13	15	14

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

Solution:

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-and-leaf plot first.

   Put the boys' scores on the left, and the girls' scores on the right. Then 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.

Unordered back-to-back stem-and-leaf plot Ordered back-to-back stem-and-leaf plot

Key: 1|2 = 12

Leaf Boys	Stem	Leaf Girls
	1	
3 2 3 3	1	2
4 5 4 5 4	1	5
6	1	6 7 6 7 6
	1	9 9 9

Key: 1|2 = 12

Leaf Boys	Stem	Leaf Girls
3 3 3 2	1	2
5 5 4 4 4	1	5
6	1	6 6 6 7 7
	1	9 9 9

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.

Exercise 1

1. 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.

   Year 12: 20, 22, 12, 8, 19, 13, 15, 23, 7, 22, 18, 20, 13, 19, 21, 16, 24, 10, 26, 24
   Year 11: 30, 6, 30, 23, 9, 20, 31, 43, 25, 41, 30, 25, 26, 30, 21, 30, 29, 15, 15, 44

   a. Complete the unordered back-to-back stem-and-leaf plot for this data.

Year 11 Leaf	Stem	Leaf **Year 12**
	0	
	1	
	2	
	3	
	4	

b. Complete the ordered back-to-back stem-and-leaf plot for this data.

Year 11 Leaf	Stem	Leaf **Year 12**
	0	
	1	
	2	
	3	
	4	

c. For each year find:

	i. the median score	ii. the range	iii. any outliers
Year 11			
Year 12			

d. State one similarity and one difference between the two sets of data.

- Similarity

- Difference

e. Mr Farley believes there is more absenteeism in Year 11 than in Year 12. What other information would we need to evaluate this statement?

2. 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.

Boys	3.4	5.0	4.2	3.7	4.9	3.4	3.8	4.8	3.6	4.3
Girls	3.0	2.7	3.7	3.3	4.0	3.1	2.6	3.2	3.6	3.1

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

Girls
Median =
Mean =

Boys
Median =
Mean =

Statement:

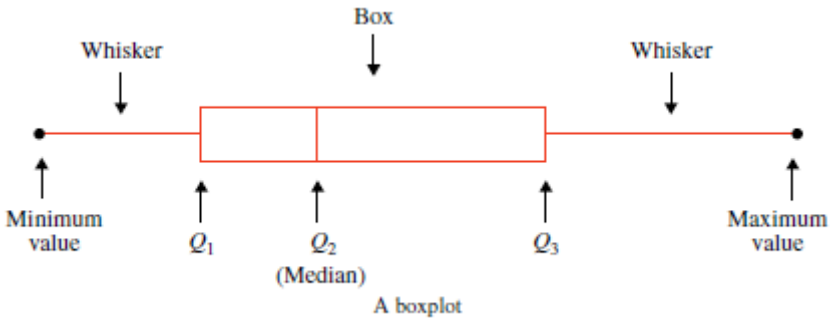
FIVE NUMBER SUMMARY AND BOXPLOTS

The five number summary statistics are as follows:

Lowest score, lower quartile Q_1, median Q_2, upper quartile Q_3, 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.

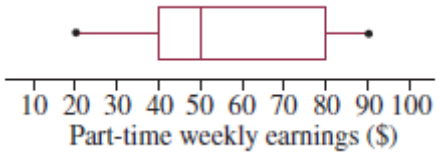
A boxplot displays the minimum and maximum values of the data together with the quartiles and is drawn with a scale. The length of the box gives us the interquartile range. A boxplot gives us a very clear visual display of how the data are spread out.



A boxplot

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.

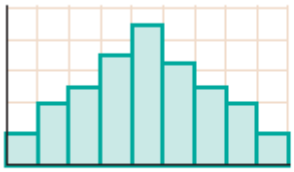


Solution:

Range = Maximum value – minimum value — Range = 90 – 20 = 70
Median is located at the bar inside the box — Median = 50
Interquartile range (IQR) = $Q_3 - Q_1$ — IQR = 80 – 40 = 40

SKEWNESS

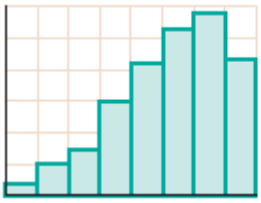
In the figure below a **symmetric distribution** is represented in the histogram and in the boxplot. The characteristics of this boxplot are that the whiskers are about the same length and the median is located about halfway along the box.



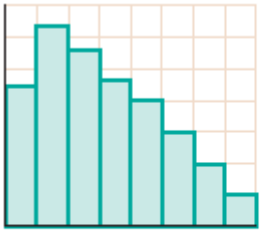
Symmetric histogram

Symmetric boxplot

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.

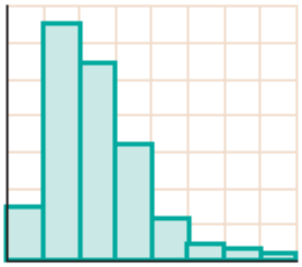


Positively skewed histogram

Positively skewed boxplot

Example:

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



Solution:

The histogram shows a distribution which is positively skewed. The boxplot shows a distribution which is approximately symmetric.

This means:

The histogram and the boxplot could not represent the same data since the histogram shows a distribution that is positively skewed and the boxplot shows a distribution that is approximately symmetric.

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

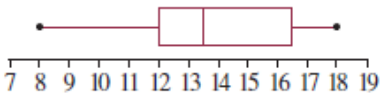
Display these data using a boxplot.

Solution:

Step 1: Order the scores from lowest to highest. Find lowest and highest score, Q_1, median, Q_3.

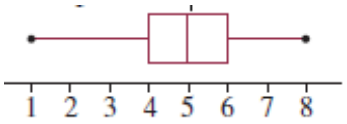
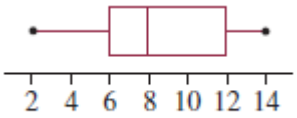
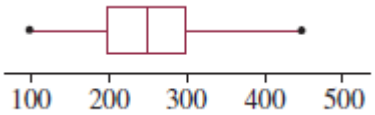
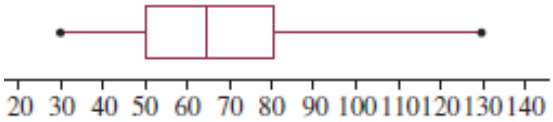
8, 11, 12, 12, 13, 13, 14, 15, 16, 17, 17, 18

Median: 13.5 **Q_1:** 12 **Q_3:** 16.5

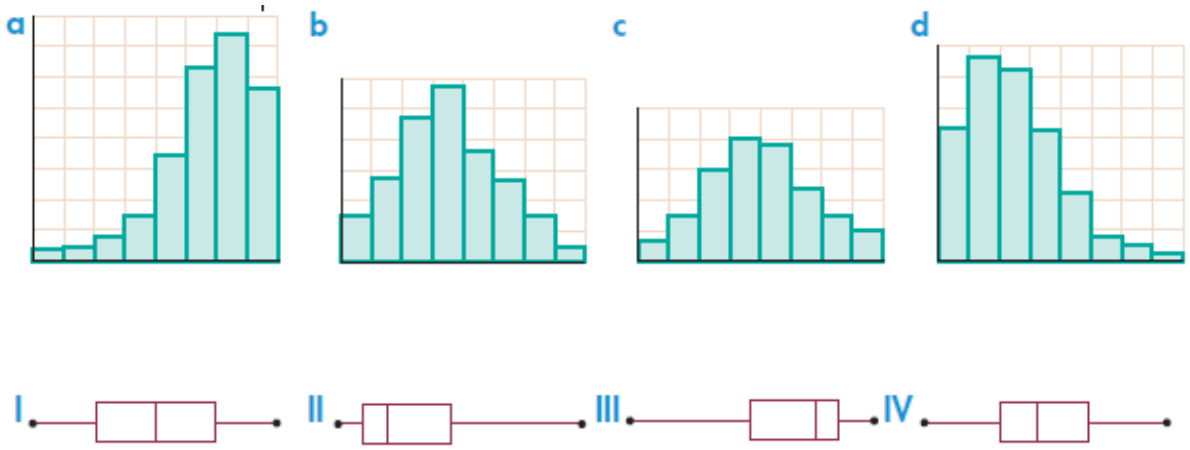


Exercise 2

1. For the boxplots shown, write down the range, interquartile range and median of the distributions.

	Range	Interquartile range	Median
			
			
			
			

2. 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.



3. 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

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

18	26	28	23	16	19	21	27	31	23	24	26
21	18	26	27	23	21	24	20	19	25	27	32
29	21	16	19	23	25	27					

   Represent this data in a boxplot. *You may use a spreadsheet to help you sort the data and find the five number summary.*

PARALLEL BOXPLOTS

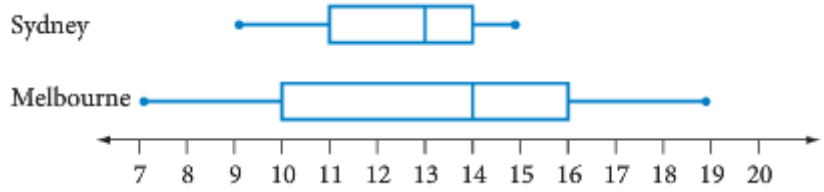
In statistics there are many opportunities to compare two sets of data. We can compare sets of data by drawing two or more boxplots using a common scale.

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 Interquartile range for Sydney = 14 – 11 = 3

Median for Melbourne = 14 Interquartile range for Melbourne = 16 – 10 = 6

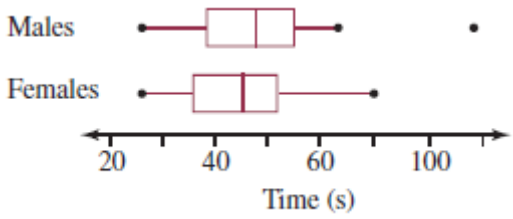
Comparison:

- 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 3

1. 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.

2. 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 term exams. Who has performed better in relation to their own class? Justify your answer.

   d. Can we calculate the mean from the information given? Explain.

3. This back-to-back stem-and-leaf plot shows the number of points scored in each match by two basketball teams during last season.

Langley Lynx		Blakely Bears
66543	4	49
8830	5	23368
8866311	6	5689
7430	7	00136
665	8	257799
22	9	034

a. How many games were played in the season?

b. Find the range for each team.

c. Find the five number summary for each team.

d. Draw a double boxplot for both teams.

e. Which is the better scoring team? Explain.

Week 5 Portfolio Task Name:

Alison has been working with some weather data, a collection of average temperatures in Fahrenheit.

Barry has also been working with data. His data set is a collection of teenagers' weights in kilograms.

They have been collecting samples of 40 data points to analyse. Unfortunately, they forgot to label their samples and can't work out which data come from which set.

Alison knows that set A came from her weather data, and Barry remembers that set B is one of his weight's samples. **Can you work out which other lists belong to Alison, and which belong to Barry?** Clearly show how you worked out which lists are average temperatures in Fahrenheit and which are teenagers' weights in kilograms.

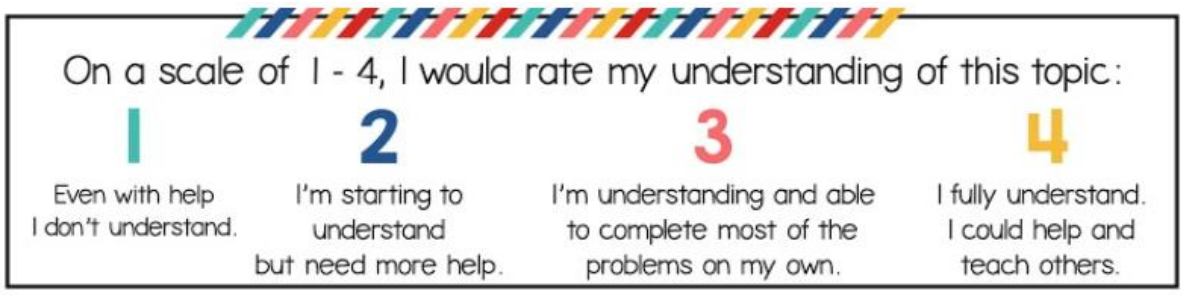
The lists appear in the table on the next page and is available as a spreadsheet on Google Classroom.

A	B	C	D	E	F
68	53	48	69	52	62
50	60	72	58	51	63
34	56	58	52	73	52
51	54	61	75	64	55
50	48	56	74	51	58
68	65	48	54	49	45
71	59	61	54	42	53
69	54	47	52	54	56
76	58	58	63	53	56
48	57	65	57	47	60
71	60	63	49	74	54
69	57	55	49	73	63
49	60	55	55	48	61
51	53	62	65	58	59
68	60	49	68	53	59
56	58	55	49	55	64
52	60	54	52	56	59
59	58	62	55	50	57
54	61	53	73	78	58
65	56	58	56	46	55
71	58	60	49	67	56
49	58	58	67	71	70
61	57	52	57	70	66
52	67	63	70	52	54
53	54	59	49	69	60
46	65	52	74	43	56
60	57	52	65	45	58
46	64	54	69	64	58
48	58	51	50	51	70
70	57	59	49	64	63
65	48	63	55	58	53
66	61	57	51	85	48
42	58	68	60	70	58
58	59	59	68	46	65
73	51	66	65	60	61
80	62	55	51	53	54
62	64	59	40	70	51
45	60	63	50	56	59
61	57	63	74	47	58
49	47	50	42	64	61

MARKING RUBRIC

CRITERIA	EXPECTATIONS	MARKS
Practical	Student completes practical work of the brief to an acceptable standard set by the teacher.	
*Completion of practical work is a **prerequisite** to submitting your portfolio task.* Your teacher reserves the right not to accept submission of your weekly/fortnightly task if the supporting work is incomplete.		
Portfolio Task	Student response is correct.	**/2**
	Student response shows clear and logical working out.	**/2**
	Student response includes appropriate units and correct rounding, where relevant.	**/2**
	Student response states a conclusion which answers the question. • *Check:* Does your answer make sense in the given context?	**/2**
	Submission Guidelines	
Timeliness	Student submits the exercises and assessable task by the set deadline.	**/2**
	TOTAL	**/10**

Student Reflection:



Written reflection (optional): What was interesting? What did you find easy? What do you need to work on? Any other comments?