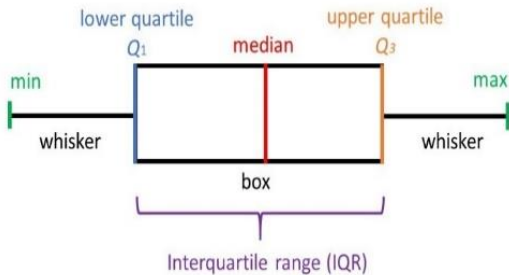# HAWKER COLLEGE
Engage | Inspire | Achieve

Goals

This week we are going to:

- construct and use parallel box plots (including the use of the 'Q1 – 1.5 x IQR' and 'Q3 + 1.5 x IQR' criteria for identifying possible outliers) to compare groups in terms of location (median), spread (IQR and range) and outliers and to interpret and communicate the differences observed in the context of the data
- compare groups on a single numerical variable using medians, means, IQRs, ranges or standard deviations, as appropriate; interpret the differences observed in the context of the data; and report the findings in a systematic and concise manner
- implement the statistical investigation process to answer questions that involve comparing the data for a numerical variable across two or more groups; for example, are Year 11 students the fittest in the school?

Theoretical Components

Resources:
PDF file: Week 4 Notes and Exercises

Knowledge Checklist
- Use of summary statistics
  - Mean
  - Median
  - Mode
  - Range
  - Interquartile range
  - Standard deviation
  - Effect of outliers
- Describing sets of data
- Comparing sets of data
- Parallel boxplots

Order
1. Read through the notes and examples
2. Work through the exercises
3. Complete the investigation at the end of the booklet.
4. Complete the reflection at the end of the booklet
5. Come and see your teacher and make sure you are up to date.

Practical Components

Work through the exercises and show the completed tasks to your teacher.

Be sure to ask for help as you need for the successful completion of all tasks.

Remember to regularly check Google Classroom for messages.

Investigation

Complete the task at the end of the booklet and submit your work for checking. 😊

QFO

MATHEMATICAL APPLICATIONS 2

WEEK 4 NOTES & EXERCISES

USE OF SUMMARY STATISTICS

So far, for data, we have looked at:

1. Measures of central tendency

The three measures of central tendency we have looked at are:

Mean – average of the data values

Median – the middle value when the data is arranged in order (50% of the data lies below the median and 50% lies above it)

Mode – the most frequent or populate data value

2. Measures of spread

Range – the highest score minus the lowest score (describing the total spread of scores)

Interquartile range (IQR) – the upper quartile minus the lower quartile (the spread of the middle 50% of the data)

Standard deviation – how data deviates from the mean

Summary statistics are very important and informative when comparing 2 or more sets of data. Comparisons can be made, such as:

- Groups A has a higher average than Group B
- Group A has a higher median value than Group B
- Group A has a smaller range than Group B
- Group A is more consistent than Group B
- Group A scores are consistently higher than Group B

Use the mean when

i the data is numerical
ii the question asks for the average
iii the data has no outliers and is not skewed.

Use the median when

i the data is numerical
ii the question asks for a middle value
iii the data has outliers or is skewed.

Use the mode when

i the data is categorical or numerical
ii the question asks for the most common feature
iii the data has outliers or is skewed.

The word **average**, as we have seen, can refer to either the mean, median or mode:

- The **mean** is the best 'average' when there are no outliers or when a distribution is fairly symmetrical.
- The **median** is the 'best average' when there are outliers, or the distribution is skewed.
- The **mode** is the 'best average' when referring to the most popular category.

Example

Read each statement and decide whether the mean, median or mode is the best measure of the 'average'. Justify your answer.

a) The average mark in a Maths test was 63.

Use the mean — all scores are included in this average.

b) Minh takes an average shoe size.

Mode — average here means most popular size.

c) Sydney's average house price is $1 300 000.

Use the median — high and low values do not affect this average.

1. Decide whether the mean, median or mode is the best measure of the average in each case. Give a reason for your choice.

    a. average height of people gathered for a family party

    b. average life of a light globe

    c. average test mark in Mathematics

    d. average daily temperature in Alice Springs in June

    e. average life span for insurance purposes

    f. men's average shirt size

    g. average number of children per family

2. The colours of the new cars sold last week at Huxley Motors was record. The results are shown in the table below.

    a. How many new cars were sold?

Colour	Black	Blue	Red	Silver	White
Frequency	4	7	7	9	12

    b. What is the mode for this data?

    c. Why isn't it possible to find the median or mean colour?

2. The following data represents the heights (cm) of students in a Year 11 class.

| 167 | 182 | 175 | 156 | 168 | 159 | 160 | 175 | 180 | 168 | 154 |
| 159 | 175 | 179 | 163 | 169 | 178 | 171 | 161 | 170 | 157 | 165 |
| 162 |

a. Order the data using a stem-and-leaf plot

b. Find each of the following and explain what it tells you about the data set:

    i)   the mean

    ii)   the median

    iii)   the mode

    iv)   the range

    v)   the interquartile range

3. The heights (in cm) of a team of basketball players are:

204   195   184   189   193   178   199

a. Find the mean height of the team, correct to one decimal place

b. Find the median height of the team and interpret

The 184 cm tall player is replaced by a 200 cm tall player

c. Recalculate the mean

d. Recalculate the median

e. Describe how and why the mean and the median have been affected by the change of players

4. Rupert's bookstore employs the following people with annual wages as shown.

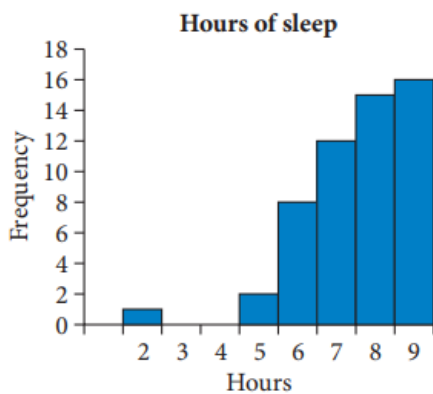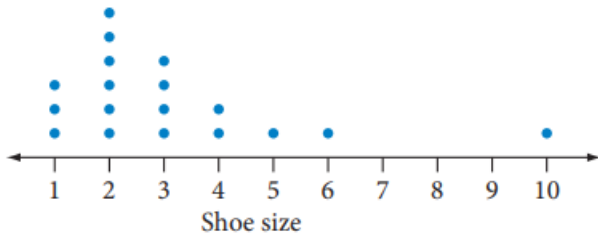|	
1 store manager	$74,300
2 cashiers	$34,200 each
2 part-time clerical staff	$28,500 each
3 salespeople	$46,500 each
2 part-time cleaners	$13,500 each

a. How many employees work for Rupert's bookstore?

b. Find the following statistical measure and explain what it tells you about the data:

   i) mean   ii) median   iii) mode

c. Which of the above measures would Rupert use to make the salaries appear higher? Why?

d. Which of the best measures best represents the annual wages for an employee at Rupert's bookstore? Why?

5. Mr. Babos owns a company, where the employees average annual salaries are being checked to ensure they are being paid fairly. The mean salary of the company is found to be $91,000, the median is found to be $80,000, while the mode is found to be $20,000. The average annual Australian salary is around $60,000.

    a. Are the employees being paid fairly? Why/why not?

    b. List a possible reason for the difference between in the mean, median and mode of the annual salaries.

    c. List 10 employee salaries that could roughly give the results above (remember, Mr. Babos is considered an employee too!)

    d. Create a boxplot to represent your data entries. Comment on the shape of the distribution.

Summary statistics	What it measures	Conditions of use
Mean	Centre	Is affected by skewed data and outliers. When investigating discrete data it may give an unrealistic answer (i.e. 2.7 children).
Median	Centre	Is not affected by skewed data and outliers.
Mode	Centre	Use when asked to find the most popular/frequent data value. (The mode is seldom used in statistics as a reliable measure of centre.)
Range	Spread	Is affected by outliers.
Interquartile range	Spread	Is not affected by outliers. ← If the data is not skewed and no outliers are present then all measures of centre and spread are appropriate.
Standard deviation	Spread	Is affected by skewed data and outliers.

Outliers need to be identified using $Q_1 - 1.5 \times IQR \leq x \leq Q_3 + 1.5 \times IQR$ and not just by visual inspection. A sentence must be written stating, for example, 43 is an outlier as it is larger than 42.

Remember the smaller the spread, the more consistent the data is.

EXERCISE 2

1. The results of fifteen students who do both Art and Music were recorded.

The mean and standard deviations were calculated as follows:

Art: mean = 43.27   sd = 15.52

Music: mean = 52.67   sd = 8.26

a. Which subject achieved the better results?

b. Which subject achieved the most consistent results?

2. For each of the following data sets, state the most appropriate measure of:

   i) centre   ii) spread
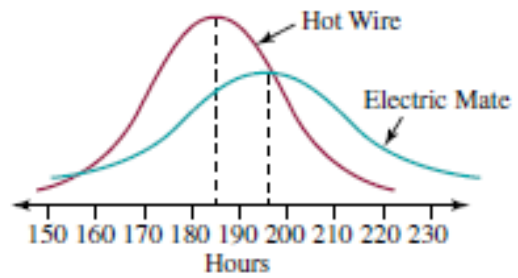
Justify each of your selections.

Shoe size

Hours of sleep

Hours

Key: 4|3 means 43%

Stem	Leaf
4	0 1 2 3 5 6 7 8 9 9 9
5	0 2 2 4 8 8 9
6	1 3 3 3 6
7	5 7 8
8	4 6
9	2

Scores

Score

3. The polygons drawn below show the lifetimes of two samples of different brands of toaster elements when subject to continued use.

a. Which brand has the longest mean life?

b. Estimate the mean life of each brand.

c. Which brand has the greatest standard deviation in its performance?

d. What does this say about the consistency of this element?

e. Which brand is best? Give a reason in support of Electric Mate. Give a reason in support of Hot Wire.

4. Crunch and Crinkle are two brands of potato crisps. Each is sold in packets nominally of the same size and for the same price. Upon investigation of a sample of packets of each it is found that Crunch and Crinkle have the same mean weight (25 g). The standard deviation of the weights of Crunch is, however, 5 g and the standard deviation of the weights of Crinkle is 2 g. Which brand do you think would represent the best value for money under these circumstances? Why?

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.
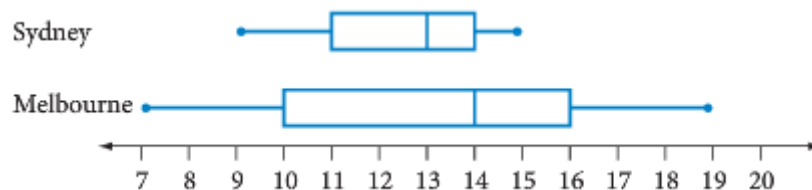
Example

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.

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.
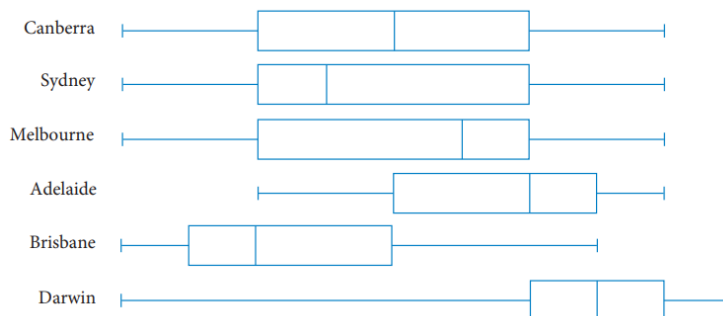

Time (s)
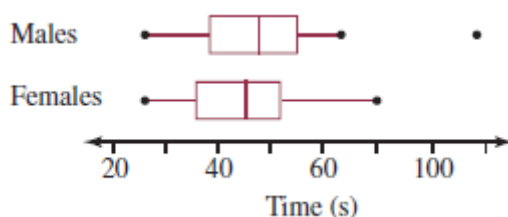
    a. Identify one similarity and one difference between the concentration spans of boys and girls.

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

2. The box plots below show the distribution of petrol prices in six Australian capital cities.

    a. Which city has:

       i)   the greatest range of prices?

       ii)   the lowest median price?

       iii) the smallest interquartile range?

       iv) Is a symmetrical distribution of prices?

    b. Is petrol cheaper in Sydney than Melbourne? How can you tell?

3. 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? Explain.

4. A survey was conducted of 130 purchasers of new black sports cars, 130 purchasers of new red sports cars, and 130 purchasers of new white sports cars. In it, people were asked the age they were when they purchased their car. The following box plots display the results:

Black sports cars

Red sports cars

White sports cars

25 30 35 40 45 50 55 60 65 70 75 80
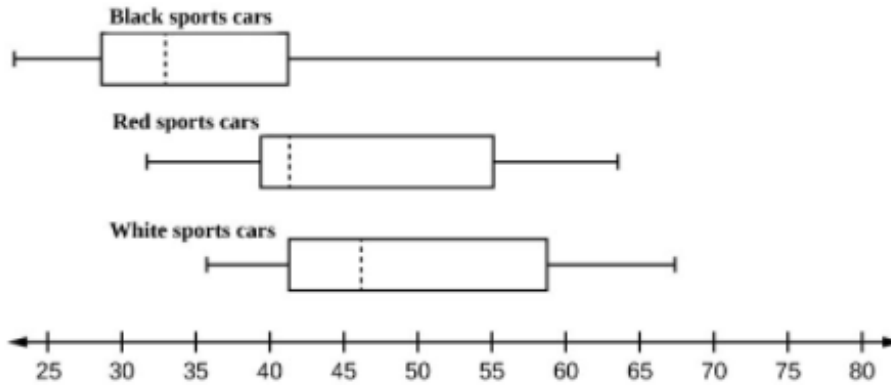
a. In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected for that car series.

Black sports cars:

Red sports cars:

White sports cars:

b. Compare the three box plots. What do they imply about the age of purchasing a sports car from the series when compared to each other?

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

Charlie 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 Charlie 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 Charlie?** Clearly show how you worked out which lists are average temperatures in 16ahrenheit 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	POSS	MULT	GIVEN	TOTAL
Practical	Student completes practical work of the brief to an acceptable standard set by the teacher.	**2**	**3**		**/6**
Investigation	Student completes the investigation of the brief to an acceptable standard set by the teacher.	**2**	**2**		**/4**
Reasoning and communications	Student responses are accurate and appropriate in presentation of mathematical ideas in different contexts, with clear and logical working out shown.	**4**	**-**		**/4**
Concepts and techniques	Student submitted work selects and applies appropriate mathematical modelling and problem solving techniques to solve practical problems, and demonstrates proficiency in the use of mathematical facts, techniques and formulae.	**4**	**-**		**/4**
	Submission Guidelines				
Timeliness	Student submits the exercises and portfolio task by the set deadline. See scoring guidelines for specific details.	**2**	**-**		**/2**
		FINAL			**/20**

Student Reflection:
How did you go with this week's work?

What was interesting?

What did you find easy?

What do you need to work on?