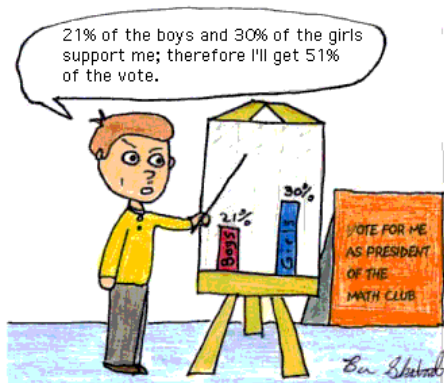


Goals

Goals for this fortnight:

- identify the mode
- calculate measures of central tendency, the arithmetic mean and the median
- investigate the suitability of measures of central tendency in various real-world contexts
- investigate the effect of outliers on the mean and the median
- calculate and interpret quartiles, deciles and percentiles
- use informal ways of describing spread, such as spread out/dispersed, tightly packed, clusters, gaps, more/less dense regions, outliers
- calculate and interpret statistical measures of spread, such as the range, interquartile range and standard deviation
- investigate real-world examples from the media illustrating inappropriate uses, or misuses, of measures of central tendency and spread



Theoretical Components

STEP 1

Resources:

PDF file: Week 3 and 4 Notes and Exercises

This Fortnight:

We will be learning:

- Mode, median and mean
- Mean, median and outliers
- Quartiles, range and interquartile range
- Deciles and percentiles
- Standard deviation
- Graphs and spread
- Using spreadsheets

Practical Components

STEP 2

Read through Week 3 and 4 Notes and Exercises for instructions on what to do.

There are 6 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

Remember to check Google Classroom or hawkermaths.com for each week's learning brief

Notes:

ESSENTIAL MATHEMATICS 2

WEEK 3 AND 4 – SUMMARISING AND INTERPRETING DATA

How are we ever going to use this?

- To calculate summary statistics to allow comparisons of data
- To understand the different 'averages' and know which is best in different situations
- To determine whether the data has been analysed correctly in order to support opinions

Statistics involves the collection of information. This is usually called data and collected in the form of numbers. This data is then organised – usually into tables and then analysed. This analysed data is used to pick out trends (e.g. is the average daily temperature rising) and importantly to make predictions (e.g. the average temperature of the earth in 50 years).

MODE, MEDIAN AND MEAN

Most people are familiar with averages. Averages tell us something about a 'typical' or 'central' value.

Statisticians have three different ways to describe the central tendencies. These ways are called the median, the mode and the mean.

- The **mode** is the most common or frequent score.
- The **median** is the middle score when the scores are placed in order from smallest to largest.
- The **mean** is calculated by adding all the scores and dividing by the number of scores. This is what most people call the 'average'

Example 1

Patricia writes the gestation period (in days) for 11 alpacas born in May. The periods are:

331 332 335 338 340 351 331 335 340 331 331

Then she puts the scores in numerical order in a stem-and-leaf plot.

Stem	Leaf
33	1 1 1 1 2 5 5 8
34	0 0
35	1

The middle score is 335.

For these scores, find the:

- a mode b median
c mean d range

Solution 1

- a Patricia works out that the mode is 331 days because there are more scores of 331 than any other score.
- b She finds that the middle score is 335 because there are five scores lower than 335 and five scores bigger than 335. The median is 335.
- c To calculate the mean, she adds up all the scores and divides the sum by 11 (because there are 11 scores).
- $$\begin{aligned}\text{Mean} &= (331 + 331 + 331 + 331 + 332 + 335 + 335 + 338 + 340 + \\ &\quad 340 + 351) \div 11 \\ &= 3695 \div 11 \\ &= 335.9 \text{ days (to 1 decimal place)}\end{aligned}$$
- d The range is $351 - 331 = 20$ days.

Example 5

Patricia records how long it takes 40 healthy baby alpacas to stand up after they are born. She records the times, correct to the nearest 10 minutes.

30 20 50 20 60 40 20 40
60 30 60 80 30 90 10 30
50 40 50 30 70 50 80 40
10 30 80 50 40 30 70 40
30 40 90 70 30 80 30 20



She uses a frequency table to organise the information.

Score (x) Time in minutes	Tally	Frequency (f)	$f \times x$
10		2	20
20		4	80
30	 	10	300
40	 	7	280
50		5	250
60		3	180
70		3	210
80		4	320
90		2	180
Total		40	1820

The tally column helps you work out how many of each score there are.

The frequency column shows the number of each score.

There are two 10s.

All the 10s add up to 20.

All the 40s add up to 280.

All the scores add up to 1820.

This number is always the total number of scores.

- Find the range, mode and mean of these times.
- How does the time it takes baby Rose to stand up compare to these times?

Solution 5

a Range = biggest score – smallest score

$$\begin{aligned} &= 90 - 10 \\ &= 80 \text{ minutes} \end{aligned}$$

Mode = 30 minutes because this was the most common time.

$$\begin{aligned} \text{Mean} &= (\text{Sum of all the scores}) \div (\text{Number of scores}) \\ &= 1820 \div 40 \\ &= 45.5 \text{ minutes} \end{aligned}$$

- Rose was born at 10:15 am and she stood up at 12:30 pm. It took Rose 2 hours and 15 minutes (i.e. 135 minutes) to stand up. Therefore Rose took almost three times the mean time to stand up. She was slow to stand up.

Exercise 1

1. Find the mode of each set of scores.

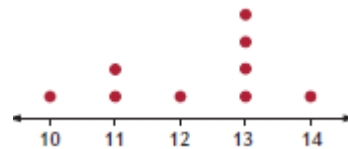
a. 9, 7, 8, 7, 10, 7

b. 16, 17, 19, 19, 20, 20, 20, 21, 23

c.

Stem	Leaf
2	6 8 8
3	1 2 2 2 8 8
4	0 3

d.



2. Arrange each set of scores from smallest to largest, and then work out the median.

a. 8, 6, 9, 5, 7

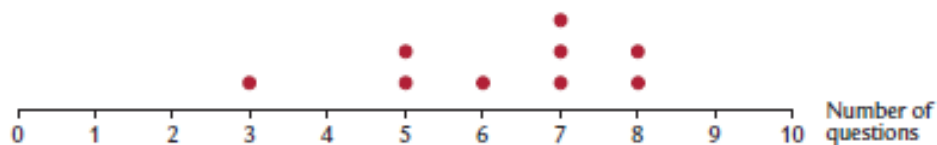
b. 11, 19, 13, 17, 16, 10, 12

3. Calculate the mean of each set of scores.

a. 28, 36, 30, 25, 21

b. 103, 6, 9, 28, 74, 92

4. This dot plot shows the number of trivia questions that the contestants at a trivia night were able to answer correctly.



a. How many contestants were there?

b. Calculate the value of the:

i. mode

ii. mean

iii. median

5. Wendy recorded the results of a spelling quiz using a stem and leaf plot.

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

Key: 3|1 is 31

- How many students sat for the spelling quiz?
- What was the lowest score?
- What is the modal score?
- Calculate the median score.
- Calculate the mean of the quiz.
- Describe the difference between the median and the mean of the spelling quiz.

6. Bob likes to go fishing. This frequency distribution table shows the number of fish he catches on each day during his 35-day fishing trip.

Score (x) Number of fish	Tally	Frequency (f)	$f \times x$
0			
1			
2			
3			
4			
5			
	Total		

- Complete the table.
- Calculate the mean number of fish Bob catches (see example on page 4). Give your answer to 1 decimal place.

7. Minka records the number of goals her soccer team scores each match.

0 2 1 2 2 3 0 1 2 1
1 1 1 0 1 4 2 3 5 0

- Construct a frequency distribution table and show the complete data.
- How many games does the team play?
- Calculate the mean number of goals the team scores each match.

MEAN, MEDIANS AND OUTLIERS

An outlier is a score much larger (or smaller) than other scores in the data set. Outliers can have a dramatic effect on the mean. The median and mode are usually unaffected.

Example: Set A: 3, 4, 5, 5, 6, 7 Mean (\bar{x}) = 5, mode = 5, median = 5
Set B: 3, 4, 5, 5, 6, 20 \bar{x} = 7.2, mode = 5, median = 5

In the second set the median and mode are more useful than the mean.

Exercise 2

1. This data shows the ages of the Binns and Thompson families.

Ages of Binns family: 19, 31, 21, 3, 6, 14, 19, 24, 11

Ages of Thompson family: 19, 31, 21, 3, 6, 14, 19, 24, 91

a. Calculate the mean age of each family. Answer correct to 1 decimal place.

b. What is the main difference between these two sets of data?

c. What effect does the difference identified in b. have on the mean?

2. Eleven houses were sold in Keswick Street over the last two years. The selling prices are listed below.

\$620,000	\$625,000	\$700,500	\$738,000	\$625,000	
\$1,800,000	\$598,000	\$612,000	\$696,500	\$720,000	\$705,000

- a. Find the median sale price for the houses.
- b. Find the mean sale price.
- c. Which measure best describes the price of houses in Keswick Street? Justify your answer.
- d. Which price is the outlier in this data?
- e. Calculate the mean of the remaining prices when this outlier is removed. Is this mean closer to the median found in part a.?

3. Mark and Steve's batting scores for six innings of cricket are shown below.

Mark: 45, 48, 53, 38, 32, 40, 51

Steve: 23, 57, 6, 125, 65, 5, 37

- a. Calculate the mean score for each player (to 1 decimal place).

- b. Which player is better if you use the mean?

- c. Find the median score for each player.

- d. Which player is better if you use the median?

- e. Which player would you rather have in your team? Justify your answer.

4. A property developer has 40 new apartments for sale. The 20 apartments on the first 5 floors are priced at \$550,000 each. The next 8 apartments on floors 6 and 7 are priced at \$590,000 and then the next 8 apartments on floors 8 and 9 are priced at \$655,000. The three apartments on the tenth floor are \$945,000 each and the penthouse apartment on the top floor is priced at \$2.3 million.
- Determine the median price for the apartments.
 - Calculate the mean price for the apartments.
 - When advertising the apartments for sale which 'average' would the developer use. Explain.
 - The developer will be speaking to potential investors in his company. What 'average' might he use to make his company look profitable? Explain.
 - Which price is the outlier?
 - Calculate the mean after removing the outlier.

Exercise 3

1. Sue and Jason work in a fast food shop. The number of hamburgers they sold each day between 12 noon and 2 pm in the month of July is recorded below.

17, 27, 28, 18, 18, 17, 19, 19, 25, 27, 17, 19, 20, 19, 21, 26,
28, 18, 19, 20, 17, 19, 23, 24, 20, 18, 17, 20, 19, 27, 28

- Arrange the data in ascending order.
 - Find the range.
 - Find each of the quartiles Q1, Q2 and Q3.
 - Calculate the interquartile range.
2. The following data shows the daily maximum temperatures (in °C) for 15 days in Cairns in July.

32, 30, 31, 32, 31, 30, 31, 31, 31, 31, 29, 25, 28, 27, 29

For this data, find:

- the range
- each of the quartiles
- the interquartile range

3. For the data shown in the stem and leaf plot find the range, median and interquartile range.

Stem	Leaf
2	3 5 5 6 7 8 9 9
3	0 2 2 3 4 6 6 7 8 8
4	2 2 4 5 6 6 6 7 9
5	0 3 3 5 6
6	2 4
7	5 9
8	2
9	7
10	
11	4

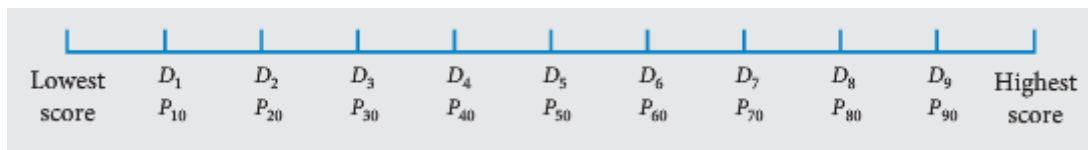
Key: 4|2 = 42

DECILES AND PERCENTILES

Another way of dividing data into groups is to divide it into deciles or percentiles. The data still must be arranged in ascending order. Deciles and percentiles are usually only used with large sets of data.

Deciles: divides the data into 10 equal groups

Percentiles: divides the data into 100 equal groups



Deciles:

D_1 cuts off the lowest 10% of scores

D_4 cuts off the lowest 40% of scores

D_9 cuts off the lowest 90% of scores or the top 10% of scores

Percentiles:

P_{24} cuts off 24% of scores.

P_{60} cuts off 60% of scores. Note P_{60} is the same as D_6

P_{87} cuts off 87% of scores, or the top 13% of scores.

Example:

The lengths (in centimetres) of 20 new-born infants at a hospital were recorded.

51 49 52 49 47 56 48 48 52 50
55 49 48 51 44 52 50 50 53 45

Place the values in order first.



- What is the 3rd decile for this data?
As D_3 lies between 48 and 49 cm the value of D_3 is 48.5 cm. Thus 30% of the babies are less than 48.5 cm long.
- What is the 5th decile for this data?
As D_5 is between 50 and 50 its value is 50 cm. This 50% of the infants are less than 50 cm or 50% of the infants are longer than 50 cm.
- What is another name for the 5th decile?
As the 5th decile is right in the middle it is also the median.
- Find the value that separates the bottom 70% from the top 30% of the infants (in terms of length).
As D_7 is between 51 and 52 cm, 51.5 cm separates the bottom 70% from the top 30% of infants.
- If the length of new-born baby James is in the top 10% of infant lengths, what value must it be greater than?
James' length must be greater than D_9 i.e. 54 cm.

Exercise 4

1. The information below is based on weather records kept by the Bureau of Meteorology. They are the maximum daily temperatures in November for Newcastle, NSW.
 - The mean is 23.5°C
 - The highest temperature on record was 41°C (on 19th November 1968)
 - The lowest temperature on record was 15.6°C (on 19th November 1986)
 - The 1st decile $D_1 = 18.9^{\circ}\text{C}$
 - The 9th decile $D_9 = 28.8^{\circ}\text{C}$
 - a. What is the range of temperatures?

 - b. What percentage of temperatures was higher than 28.8°C ?

 - c. What value would you expect the median to be close to (but not necessarily equal)? **Hint:** mode, mean, range.

 - d. What is the size of the 9th decile band (the difference between the highest temperature and the 9th decile)?

2. The table shows the percentiles for the heights (in cm) of girls aged 2 to 5 years, according to the child growth standards of the World Health Organisation (WHO).

Age (years)	P ₅	P ₂₅	P ₅₀	P ₇₅	P ₈₅	P ₉₉
2	80.4	83.5	85.7	87.9	89.1	93.2
2.5	84.9	88.3	90.7	93.1	94.3	98.9
3	88.8	92.5	95.1	97.6	99.0	103.9
3.5	92.4	96.3	99.0	101.8	103.3	108.5
4	95.6	99.8	102.7	105.6	107.2	112.8
4.5	98.7	103.1	106.2	109.2	110.9	116.7
5	101.6	106.2	109.4	112.6	114.4	120.5

- a. What is the median height of a 4 year old girl?
- b. Libby is aged 2.5 and is 88.3 cm tall. Is she tall for her age? What percentage of girls her age are shorter than her?
- c. What is Libby's expected height when she turns 5 years old?
- d. Only 15% of girls of Renee's age are taller than her. How tall is she if she is 3.5 years old?
- e. Mikayla is 2 years old and 93.2 cm tall. Is she short for her age? What percentage of girls her age are taller than her?

DESCRIBING SPREAD

The **standard deviation** is another measure of spread, like the range and the interquartile range. It describes how far scores are from the mean. The bigger the standard deviation, the more spread out the scores are. The formula for standard deviation is quite complicated so you don't need to remember it. We will use spreadsheets to find the standard deviation.

A large standard deviation indicates that the scores are more spread out, while a small standard deviation indicates scores that are closer together, or more consistent.

Example:

The net weekly wages of 8 labourers are:

\$730 \$490 \$600 \$440 \$490 \$370 \$700 \$580

Mean = \$550

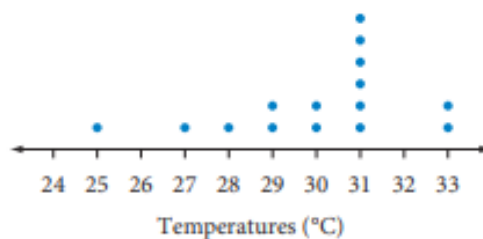
Standard Deviation = \$117.26

A large standard deviation means the wages are quite spread out.

When we display data in graphical form, for example as a dot plot, we can make judgements about the spread of scores. Scores can be **spread out** or **clustered** around one place or there might be **gaps** in the spread of scores.

Example:

Comment on the spread of the scores shown in this dot plot for temperature (°C) in Cairns in July.



Solution:

Look for clusters.

These scores are clustered around 31°.

Look for gaps.

There are gaps at the bottom end of the data and at the top end.

The temperatures range from 25 to 33.

These temperatures are quite spread out.

Write a description.

The temperatures are quite spread out but they are clustered around 31°. There are gaps towards the bottom end of the data.

Exercise 5

1. The cricket selectors are trying to choose between two pairs of indoor cricket batsmen for the state team. Both pairs have an average (mean) of 33 runs. Two sets of results for the pairs of batsmen (in runs) are:

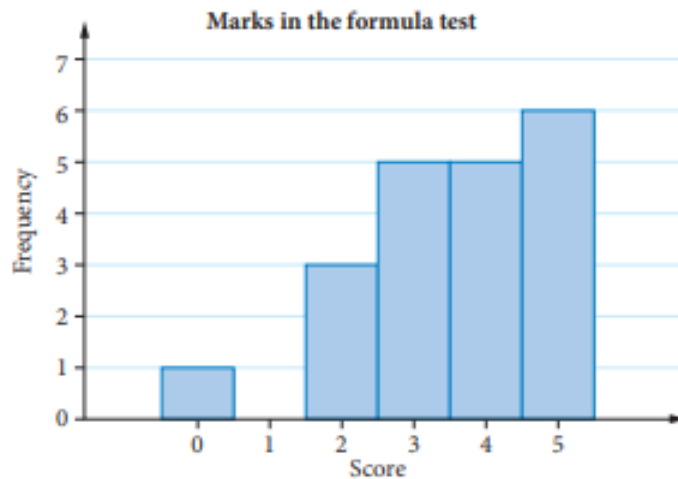
Pair A: 34, 30, 36, 35, 29, 34

Pair B: 41, 26, 37, 35, 25, 34

The standard deviation for Pair A is 2.58 runs, while the standard deviation for Pair B is 5.74 runs.

Which pair is more consistent and why?

2. Ms Thomson gave her Year 12 class a short formula test. This frequency histogram shows the results.



Comment on the spread, clusters and gaps of the scores shown in this histogram.

USING SPREADSHEETS

We can calculate the mode, mean, median and other statistics using the statistical functions on a spreadsheet.

Exercise 6

This exercise is set as an assignment on Google Classroom (available in week 4). Use the assigned Google Sheet to complete the following questions. Submit via the assignment on Google Classroom.

1.

- a. Enter into a spreadsheet the following data about the daily maximum temperatures in Alice Springs in one week.

	A	B	C	D	E
1	Day	Temperature (°C)			
2	Sunday	29		Mean	
3	Monday	31		Mode	
4	Tuesday	30		Median	
5	Wednesday	33		Standard deviation	
6	Thursday	29		Maximum	
7	Friday	28		Minimum	
8	Saturday	35		Range	
9					
10					

- b. Enter the following formulas into the given cells to find the statistics:

To find the mean, in Cell E2: **=average(B2:B8)**

To find the mode, in Cell E3: **=mode(B2:B8)**

To find the median, in Cell E4: **=median(B2:B8)**

To find the standard deviation, in Cell E5: **=stdev(B2:B8)**

To find the maximum, in Cell E6: **=max(B2:B8)**

To find the minimum, in Cell E7: **=min(B2:B8)**

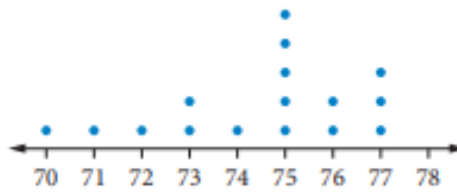
To find the range, subtract the minimum from the maximum, in Cell E8:
=E6-E7

2. Martin is looking for a job in the construction industry. He collected information on the wages of apprentices at a large construction company. The following amounts are the weekly wages of 10 apprentices.

\$342, \$528, \$684, \$401, \$574, \$386, \$601, \$475, \$558, \$427

- Find the mean wage.
- Find the standard deviation of this data. Answer correct to 1 decimal place.
- Give a reason for the large standard deviation of this data.

3. Paul's golf scores for 16 rounds of golf are shown on this dot plot.



- From the dot plot, write out a list of Paul's scores for the 16 rounds of golf.
- Calculate the mean and the standard deviation for this data.
- Is Paul a consistent player? Explain your answer.

4. The percentage scores of a class of 30 students in a science test are shown.

61	75	46	78	81	95	67	61	50	74
100	57	83	64	69	95	85	89	66	45
70	87	84	80	63	92	64	75	97	60

- Enter the data onto your spreadsheet and use the "sort" function to arrange the data in ascending order.
- What is the 8th decile?
- What is the 3rd decile?
- What is the 40th percentile?
- Find the value that cuts off the lower 20% of scores from the upper 80%.
- What percentage of students scored higher than 79?

Task 1

Find sets of integers that satisfy the following:

1. Three numbers with mean 3 and mode 2 _____
2. Three numbers with mean 7 and mode 10 _____
3. Three numbers with mean 8, median 10 and range 8 _____
4. Four numbers with mean 7.5, mode 6 and median 7

5. Four numbers with mean 6, median 6.5 and range 11

6. Five numbers with mean 4, mode 3 and range 9

7. Five numbers with mean 4, mode 2 and range 6

8. Five numbers with mean 7, mode 7 and range 10

Task 2

The three measures of central tendency are mean, median and mode.


The statistical measures of spread are range, interquartile range and standard deviation.

- Write a brief description of a real-world example in which one measure of central tendency provides useful information while another provides meaningless (or even misleading) information.
- Explain which of the three measures of central tendency would be appropriate for your example, which would be inappropriate, and why.
- Then identify which measure of spread is most suited to your example. Explain why.

MARKING RUBRIC

CRITERIA	EXPECTATIONS	MARKS
Practical	Student completes practical work of the brief to an acceptable standard set by the teacher.	
<p><i>Completion of practical work is a prerequisite to submitting your investigation.</i></p> <p>Your teacher reserves the right not to accept submission of your weekly/fortnightly investigation if the supporting work is incomplete.</p>		
Portfolio Task	Student response is correct.	<i>/2</i>
	Student response shows clear and logical working out.	<i>/2</i>
	Student response includes appropriate units and correct rounding, where relevant.	<i>/2</i>
	Student response states a conclusion which answers the question. <ul style="list-style-type: none"> • <i>Check:</i> Does your answer make sense in the given context? 	<i>/2</i>
	Submission Guidelines	
Timeliness	Student submits the exercises and investigation by the set deadline.	<i>/2</i>
	TOTAL	<i>/10</i>

Student Reflection:



On a scale of 1 - 4, I would rate my understanding of this topic:

1	2	3	4
Even with help I don't understand.	I'm starting to understand but need more help.	I'm understanding and able to complete most of the problems on my own.	I fully understand. I could help and teach others.

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