

---

# Stats1 Chapter 2: Measures of Data

## Central Tendency

# This Chapter Overview

This is identical to the equivalent chapter in the old S1 module. Some content will be familiar from GCSE (mean of grouped/ungrouped data), but many concepts new (e.g. standard deviation) along with possibly unfamiliar notation.

## 1:: Mean, Median, Mode

“Calculate the mean of this grouped frequency table.”

## 2:: Quartiles, Percentiles, Deciles

“Use linear interpolation to estimate the interquartile range.”

## 3:: Variance & Standard Deviation

“Calculate the standard deviation of the maths marks.”

## 4:: Coding

“The marks  $x$  were coded using  $y = 2x + 10$ . Given that the standard deviation of  $y$  is 5, determine the standard deviation of  $x$ .”

# Variables in algebra vs stats

$x$

## Similarities

- ❑ Just like in algebra, variables in stats **represent the value of some quantity**, e.g. shoe size, height, colour.
- ❑ As we saw in the previous chapter, variables can be discrete or continuous.
- ❑ **Can be part of further calculations**, e.g. if  $x$  represents height, then  $2x$  represents twice people's height. In stats this is known as '**coding**', which we'll cover later.

## Differences

- ❑ Unlike algebra, a variable in stats represents the value of **multiple objects** (i.e. it's a bit like a set). e.g. the heights of **all** people in a room.
- ❑ Because of this, we can do **operations** on it as if it was a **collection of values**:

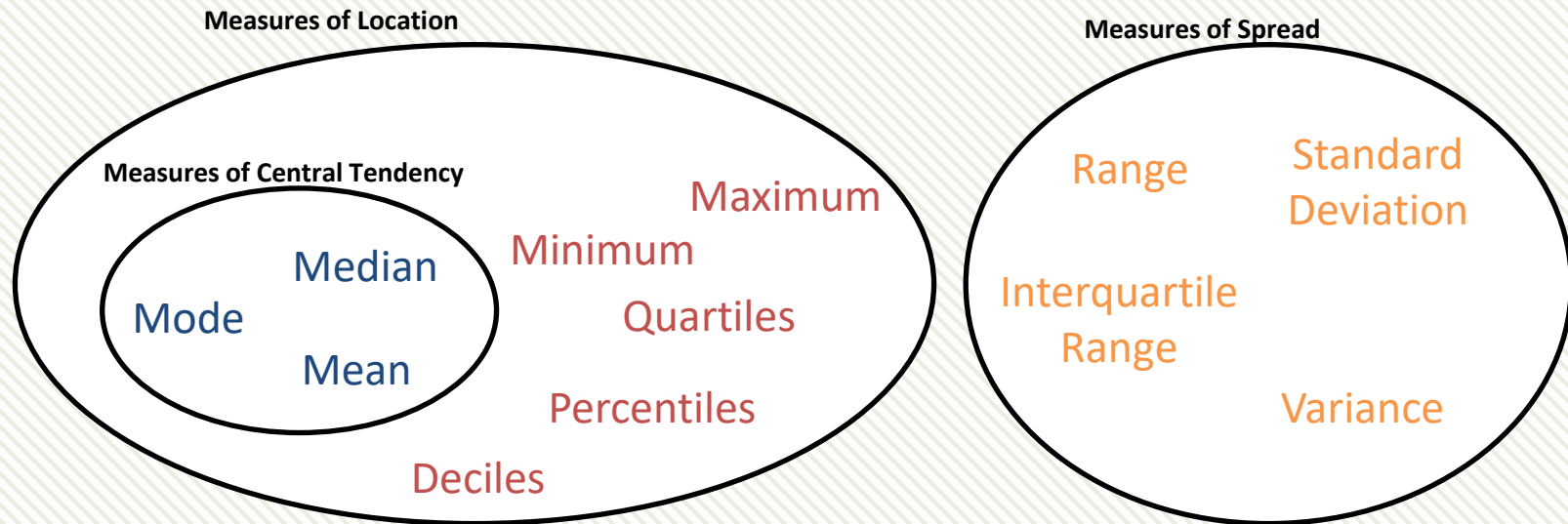
- ❑ If  $x$  represents people's heights,

$$\Sigma x$$

gives the sum of everyone's heights. In algebra this would be meaningless: if  $x = 4$ , then  $\Sigma x$  makes no sense!

- ❑  $\bar{x}$  is the mean of  $x$ . Notice  $x$  is a collection of values whereas  $\bar{x}$  is a single value.
- ❑ To each value of the variable, **we could attach an associated probability**. This is known as a **random variable** (Chapter 6).

# Measures of ...



**Measures of location** are single values which describe a **position** in a data set.

Of these, **measures of central tendency** are to do with the **centre of the data**, i.e. a notion of 'average'.

**Measures of spread** are to do with **how data is spread out**.

# Mean of ungrouped data

Diameter of coin $x$ (cm)	2.2	2.5	2.6	2.65	2.9
------------------------------	-----	-----	-----	------	-----

You all know how to find the mean of a list of values. But lets consider the notation, and see how theoretically we could calculate each of the individual components on a calculator.

$$\bar{x} = \frac{\Sigma x}{n}$$

The 'overbar' in stats specifically means 'the sample mean of', but don't worry about the 'sample' bit for now.



"Use of Technology" Monkey says:  
Time to whip out yer Casios...

# Inputting Data

Diameter of coin $x$ (cm)	2.2	2.5	2.6	2.65	2.9
------------------------------	-----	-----	-----	------	-----

Use the MENU button to access STATS mode.

## On older black/silver Casios:

- Select 1-VAR meaning “1 variable”.
- Enter each value above, pressing = after each entry.
- Press AC to start a statistical calculation.
- Use SHIFT → 1 to access statistical symbols, e.g.  $\bar{x}$ , to insert into your calculation. Press = to evaluate when done.

$$n = 5$$

$$\Sigma x / n = 2.57$$

$$3\bar{x} + 1 = 8.71$$

$$\Sigma x = 12.85$$

$$\bar{x} = 2.57$$

## On a Classwiz:

- Select 1-Variable.
- Enter each value above, pressing = after each entry.
- Press AC to start a statistical calculation.
- Press the OPTN button. “1-Variable Calc” will calculate all common statistics (including all on the left). Alternatively you can construct a statistical expression yourself – in the OPTN menu press Down. “Variable” for example contains  $\bar{x}$ . This will insert it into your calculation; press = when done.

# Frequency Tables (ungrouped data)



Number of Children ( $x$ )	Frequency ( $f$ )
0	4
1	3
2	9
3	2

A frequency table allows us to avoid writing out duplicated values. Recall that each value  $x$  must be multiplied by the frequency ( $f$ ), to ensure each value is duplicated appropriately when adding up all the values.

Mean: 
$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{27}{18} = 1.5$$

**Exam Tip:** In the exam you get a method mark for the division and an accuracy mark for the final answer. Write:

$$\bar{x} = \frac{46.75}{40} = 1.16875$$

You're not required to show working like " $0 \times 4 + \dots$ "

# Doing it in STATS mode



Number of Children ( $x$ )	Frequency ( $f$ )
0	4
1	3
2	9
3	2

To add a frequency column for data input, press SHIFT → SETUP, press Down, then choose Statistics. Turn frequency 'On'.

You can then input data in the usual way. Use the arrows to scroll back to the top of the table

How on the calculator  
would we get...

$$\begin{array}{l} \nearrow \Sigma f x \\ \searrow \Sigma f \end{array}$$

Just use  $\Sigma x$ . Your calculator doesn't have a concept of the variable  $f$  – it effectively just duplicates the values so that it's doing the calculations on normal listed data.

This is  $n$ . (Since  $n = \Sigma f$ )

**Opinion:** I wouldn't even bother remembering the  $\bar{x} = \frac{\Sigma f x}{\Sigma f}$  formula.

I'd instead just remember  $\bar{x} = \frac{\Sigma x}{n}$  and just think what the  $\Sigma x$  and  $n$  mean in the context of frequency tables. A further justification to ignore it is due to the discussion above; your calculator has no concept of the variable  $f$ .

**Quartiles!** The ClassWiz (but not older models) will calculate quartiles ( $Q_1$  is lower quartile,  $Q_3$  upper). This will be in the list of statistics when you use "1-Variable Calc". You can also insert them into your calculation by finding them under the "Min/Max" menu. However, this is not applicable if your data was grouped.



# Grouped Data



Height $h$ of bear (in metres)	Frequency
$0 \leq h < 0.5$	4
$0.5 \leq h < 1.2$	20
$1.2 \leq h < 1.5$	5
$1.5 \leq h < 2.5$	11

We don't know the exact values anymore. So what do we assume each value is?

?

Estimate of Mean:

$\bar{x} =$

?

$=$

?

$=$

?

Why is our mean just an estimate?

?

**Warning:** ClassWiz will calculate the lower and upper quartiles ( $Q_1$ ,  $Q_3$ ) along with the median. However, this is not applicable to grouped data: When you input your midpoints in the data input, your calculator doesn't know these are midpoints – it just assumes for example that the first 4 bears did have a height of 0.25m. We need to take into account the class widths to estimate the median and quartiles (which we'll see later), and your calculator cannot do this.

# Grouped Data



Height $h$ of bear (in metres)	Frequency
$0 \leq h < 0.5$	4
$0.5 \leq h < 1.2$	20
$1.2 \leq h < 1.5$	5
$1.5 \leq h < 2.5$	11

We don't know the exact values anymore. So what do we assume each value is?  
**The midpoint of each interval. We use the variable  $x$  to indicate the midpoint.**  
**We can then calculate mean in exactly the same way as before.**

Estimate of Mean: 
$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{46.75}{40} = 1.17m$$

Why is our mean just an estimate?

Because we don't know the exact heights within each group. Grouping data loses information.

**Warning:** ClassWizs will calculate the lower and upper quartiles ( $Q_1$ ,  $Q_3$ ) along with the median. However, this is not applicable to grouped data: When you input your midpoints in the data input, your calculator doesn't know these are midpoints – it just assumes for example that the first 4 bears did have a height of 0.25m. We need to take into account the class widths to estimate the median and quartiles (which we'll see later), and your calculator cannot do this.

# Mini-Exercise

Use your calculator's STATS mode to determine the mean (or estimate of the mean).

**Ensure that you show the division in your working.**

1

Num children ( $c$ )	Frequency ( $f$ )
0	2
1	6
2	1
3	1

$\bar{c} =$

?

2

IQ of L6Ms2 ( $q$ )	Frequency ( $f$ )
$80 < q \leq 90$	7
$90 \leq q < 100$	5
$100 \leq q < 120$	3
$120 \leq q < 200$	1

$\bar{x} =$

?

3

Time $t$	Frequency ( $f$ )
$9.5 < t \leq 10$	32
$10 \leq t < 12$	27
$12 \leq t < 15$	47
$15 \leq t < 16$	11

$\bar{x} =$

?

# Mini-Exercise

Use your calculator's STATS mode to determine the mean (or estimate of the mean).

**Ensure that you show the division in your working.**

1

Num children ( $c$ )	Frequency ( $f$ )
0	2
1	6
2	1
3	1

$$\bar{c} = \frac{\Sigma c}{n} = \frac{11}{10} = 1.1$$

2

IQ of L6Ms2 ( $q$ )	Frequency ( $f$ )
$80 < q \leq 90$	7
$90 \leq q < 100$	5
$100 \leq q < 120$	3
$120 \leq q < 200$	1

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1560}{16} = 97.5$$

3

Time $t$	Frequency ( $f$ )
$9.5 < t \leq 10$	32
$10 \leq t < 12$	27
$12 \leq t < 15$	47
$15 \leq t < 16$	11

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1414}{117} = 12.1 \text{ to 3sf.}$$

# GCSE RECAP :: Combined Mean

The mean maths score of 20 pupils in class A is 62.

The mean maths score of 30 pupils in class B is 75.

- What is the overall mean of all the pupils' marks.
- The teacher realises they mismarked one student's paper; he should have received 100 instead of 95. Explain the effect on the mean and median.

This subtopic doesn't appear in your textbook but has cropped up in exams.

?

## Test Your Understanding

Archie the Archer competes in a competition with 50 rounds. He scored an average of 35 points in the first 10 rounds and an average of 25 in the remaining rounds. What was his average score per round?

?

# GCSE RECAP :: Combined Mean

This subtopic doesn't appear in your textbook but has cropped up in exams.

The mean maths score of 20 pupils in class A is 62.

The mean maths score of 30 pupils in class B is 75.

- What is the overall mean of all the pupils' marks.
- The teacher realises they mismarked one student's paper; he should have received 100 instead of 95. Explain the effect on the mean and median.

$$\text{Mean} = \frac{(20 \times 62) + (30 \times 75)}{50} = \frac{3490}{50} = 69.8$$

**The revised score will increase the mean but leave the median unaffected.**

## Test Your Understanding

Archie the Archer competes in a competition with 50 rounds. He scored an average of 35 points in the first 10 rounds and an average of 25 in the remaining rounds. What was his average score per round?

$$\text{Mean} = \frac{(35 \times 10) + (25 \times 40)}{50} = 27$$

# Median – which item?

You need to be able to find the median of both listed data and of grouped data.

## Listed data

Items	$n$	Position of median	Median
1,4,7,9,10	5	?	?
4,9,10,15	4	?	?
2,4,5,7,8,9,11	7	?	?
1,2,3,5,6,9,9,10,11,12	10	?	?

Can you think of a rule to find the position of the median given  $n$ ?




## Grouped data

IQ of L6Ms2 ( $q$ )	Frequency ( $f$ )
$80 \leq q < 90$	7
$90 \leq q < 100$	5
$100 \leq q < 120$	3
$120 \leq q < 200$	2

Position to use for median:

?

 To find the median of grouped data, find  $\frac{n}{2}$ , then use linear interpolation.

**DO NOT** round  $\frac{n}{2}$  or adjust it in any way.

This is just like at GCSE where, if you had a cumulative frequency graph with 60 items, you'd look across the 30<sup>th</sup>. We'll cover linear interpolation in a sec...

# Median – which item?

You need to be able to find the median of both listed data and of grouped data.

## Listed data

Items	$n$	Position of median	Median
1,4,7,9,10	5	3 <sup>rd</sup>	7
4,9,10,15	4	2 <sup>nd</sup> /3 <sup>rd</sup>	9.5
2,4,5,7,8,9,11	7	4 <sup>th</sup>	7
1,2,3,5,6,9,9,10,11,12	10	5 <sup>th</sup> /6 <sup>th</sup>	7.5

Can you think of a rule to find the position of the median given  $n$ ?



To find the position of the median for listed data, find  $\frac{n}{2}$ :

- If a decimal, round up.
- If whole, use halfway between this item and the one after.

## Grouped data

IQ of L6Ms2 ( $q$ )	Frequency ( $f$ )
$80 \leq q < 90$	7
$90 \leq q < 100$	5
$100 \leq q < 120$	3
$120 \leq q < 200$	2

Position to use for median:

**8.5**



To find the median of grouped data, find  $\frac{n}{2}$ , then use linear interpolation.

**DO NOT** round  $\frac{n}{2}$  or adjust it in any way.

This is just like at GCSE where, if you had a cumulative frequency graph with 60 items, you'd look across the 30<sup>th</sup>. We'll cover linear interpolation in a sec...



# Exercises 2.1

Pearson Statistics & Mechanics Year 1/AS

Pages 8-9

---

# Homework Exercise

- 1 Meryl collected wild mushrooms every day for a week. When she got home each day she weighed them to the nearest 100 g. The weights are shown below:

500    700    400    300    900    700    700

- a Write down the mode for this data.
- b Calculate the mean for this data.
- c Find the median for this data.

**Hint** Try to answer part **d** without recalculating the averages. You could recalculate to check your answer.

On the next day, Meryl collects 650 g of wild mushrooms.

- d Write down the effect this will have on the mean, the mode and the median.
- 2 Joe collects six pieces of data,  $x_1, x_2, x_3, x_4, x_5$  and  $x_6$ . He works out that  $\Sigma x$  is 256.2.
- a Calculate the mean for this data.
- He collects another piece of data. It is 52.
- b Write down the effect this piece of data will have on the mean.

- 3 From the large data set, the daily mean visibility,  $v$  metres, for Leeming in May and June 2015 was recorded each day. The data is summarised as follows:

May:  $n = 31, \Sigma v = 724\,000$

June:  $n = 30, \Sigma v = 632\,000$

- a Calculate the mean visibility in each month.
- b Calculate the mean visibility for the total recording period.

**Hint** You don't need to refer to the actual large data set. All the data you need is given with the question.

- 4 A small workshop records how long it takes, in minutes, for each of their workers to make a certain item. The times are shown in the table.

Worker	A	B	C	D	E	F	G	H	I	J
Time in minutes	7	12	10	8	6	8	5	26	11	9

- a Write down the mode for this data.
- b Calculate the mean for this data.
- c Find the median for this data.
- d The manager wants to give the workers an idea of the average time they took. Write down, with a reason, which of the answers to **a**, **b** and **c** she should use.

# Homework Exercise

- 5 The frequency table shows the number of breakdowns,  $b$ , per month recorded by a road haulage firm over a certain period of time.

Breakdowns	0	1	2	3	4	5
Frequency	8	11	12	3	1	1

- Write down the modal number of breakdowns.
- Find the median number of breakdowns.
- Calculate the mean number of breakdowns.
- In a brochure about how many loads reach their destination on time, the firm quotes one of the answers to **a**, **b** or **c** as the number of breakdowns per month for its vehicles. Write down which of the three answers the firm should quote in the brochure.

- 6 The table shows the frequency distribution for the number of petals in the flowers of a group of celandines.

Number of petals	5	6	7	8	9
Frequency	8	57	29	3	1

Calculate the mean number of petals.

- 7 A naturalist is investigating how many eggs the endangered kakapo bird lays in each brood cycle. The results are given in this frequency table.

Number of eggs	1	2	3
Frequency	7	$p$	2

If the mean number of eggs is 1.5, find the value of  $p$ .

## Problem-solving

Use the formula for the mean of an ungrouped frequency table to write an equation involving  $p$ .

# Homework Answers

- 1 **a** 700 g                      **b** 600 g                      **c** 700 g  
   **d** The mean will increase; the mode will remain  
     unchanged; the median will decrease.
- 2 **a** 42.7  
   **b** The mean will increase.
- 3 **a** May: 23 355 m, June: 21 067 m  
   **b** 22 230 m
- 4 **a** 8 minutes              **b** 10.2 minutes    **c** 8.5 minutes  
   **d** The median would be best. The mean is affected by  
     the extreme value 26.
- 5 **a** 2                      **b** 1                      **c** 1.47              **d** the median
- 6 6.31 petals
- 7 1