# Welcome MYP 9 Mathematics!

| | Assignment Effort Grade (Circle One) | Comments (What was interesting or challenging?) |
|---|---|---|
| **Monday**<br>Date: 4/09<br>Topic: 10A: Types of Data | 0 1 2 | |
| **Tuesday**<br>Date: 4/10<br>Topic: 10B: Stem plots, discrete data | 0 1 2 | |
| **Wednesday**<br>Date: 4/11<br>Topic: 10C: Histograms, continuous data | 0 1 2 | |
| **Thursday**<br>Date: ______<br>Topic: ____________ | 0 1 2 | |
| **Friday**<br>Date: ______<br>Topic: ____________ | 0 1 2 | |

# Warm-up:

What can we say about Bo's (Ms. Berg's nephew) naps?

Length of Naps:

| | |
|---|---|
| 120 min | 80 min |
| 150 min | 120 min |
| 125 min | 90 min |
| 100 min | 125 min |
| 67 min | 120 min |
| 120 min | 120 min |
| 122 min | 20 min |

Class Plan

1. Warm-up

2. Which Measure is best?

D MEASURING THE CENTRE OF A DATA SET

3. Practice

## D MEASURING THE CENTRE OF A DATA SET

# What is a Measure of Central Tendancy?

In this course we consider two statistics that are commonly used to measure the **centre** of a data set. These are the **mean** and the **median**.

*https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php*

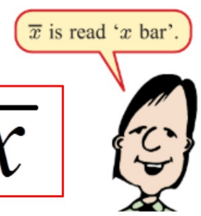
## D MEASURING THE CENTRE OF A DATA SET

### THE MEAN (Average)

The **mean** $\overline{x}$ of a data set is the statistical name for its *arithmetic average*. It can be found by dividing the sum of the data values by the number of data values.

$$\textbf{mean} = \frac{\textbf{the sum of the data values}}{\textbf{the number of data values}}$$

$$\overline{x}$$




### THE MEDIAN (Middle)

The **median** is the *middle value* of an ordered data set.

For an **odd number** of data, the median is one of the data.

For an **even number** of data, the median is the average of the two middle values. The median might not be one of the original data.

If there are $n$ data values, find the value of $\frac{n+1}{2}$.

The median is the $\left(\frac{n+1}{2}\right)$th data value.

**THE MODE:** Most frequent value

**INVESTIGATION** **THE EFFECT OF OUTLIERS**

Data Set: 4,5,6,6,6,7,7,8,9,10

Do: Investigation in your notebook

1. Calculate mean, median, mode
2. Add extreme value/ outlier of 100. Then, calculate mean, median, mode
3. Comment on the effect that the outlier has on the mean, median, and mode.
4. Which measure of central tendancy was most affected by the outlier?

Data Set: 4,5,6,6,6,7,7,8,9,10

1. Calculate mean, median, mode

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{4+5+6+6+6+7+7+8+9+10}{10} = \frac{68}{10}$$

$$\boxed{\bar{X} = 6.8}$$

$$\text{Median} = \frac{6+7}{2} = \frac{13}{2} = \boxed{6.5}$$

$$\text{mode} = \boxed{6}$$

Data Set: 4,5,6,6,6,7,7,8,9,10

2. Add extreme value/ outlier of 100. Then, calculate mean, median, mode

OUTLIER

$$\bar{x} = \frac{4+5+6+6+6+7+7+8+9+10+100}{11} = \frac{168}{11} \approx 15.3$$

$\bar{x} \approx 15.3$

MEDIAN = 7 (Odd set of data)

mode = 6

3. Comment on the effect that the outlier has on the mean, median, and mode.

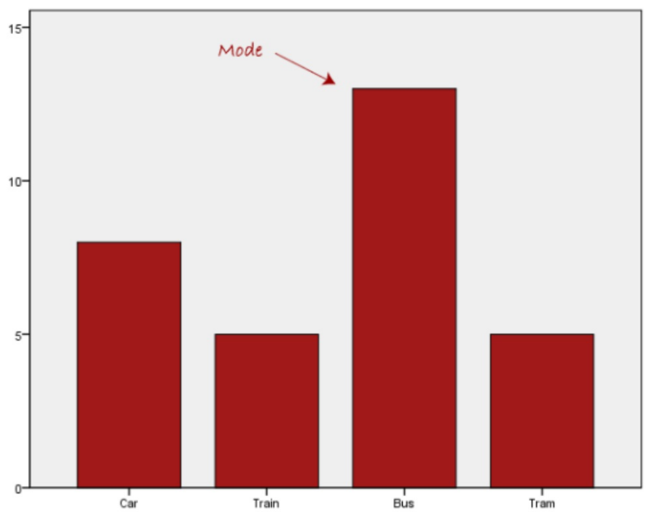
The mean was skewed by an outlier,the median and mode were not changed as drastically.

4. Which measure of central tendancy was most affected by the outlier?

The mean is skewed by an outlier, then median is the best measure of the center.

# When is the MODE the best measure of the center?

Normally, the mode is used for categorical data where we wish to know which is the most common category, as illustrated below:




Categorical data!

# Which measure of center is best?

Categorical - Mode
Numerical (skewed) - Median
Numerical (not skewed) - Mean

# Example: Maintaining an Average

Southwest made it to the playoffs. Charles Brown rushed for 78 yards, 95 yards and 64 yards in the first three games.

If Charles wants to average (mean) 80 yards, how many does he need in the final game?

Show work algebraically

# Example: Maintaining an Average

Southwest made it to the playoffs. Charles Brown rushed for 78 yards, 95 yards and 64 yards in the first three games. Show work algebraically

If Charles wants to average (mean) 80 yards, how many does he need in the final game?

4 games, $\bar{X} = 80$ yds

$$\frac{x+78+95+64}{4} = 80$$

$$\frac{x+237}{4} = 80$$

$x + 237 = 320$

$x = 83$ yards

**Example 8** — Self Tutor

Solve the following problems:

**a** The mean of six scores is 78.5. What is the sum of the scores?

**b** Find $x$ if 10, 7, 3, 6, and $x$ have a mean of 8.

**a**
$$\frac{\text{sum}}{6} = 78.5$$
$$\therefore \text{sum} = 78.5 \times 6$$
$$= 471$$
$\therefore$ the sum of the scores is 471.

**b** There are 5 scores.
$$\therefore \frac{10+7+3+6+x}{5} = 8$$
$$\therefore \frac{26+x}{5} = 8$$
$$\therefore 26 + x = 40$$
$$\therefore x = 14$$

# Ace in Tennis! - What's the mean?

**Example 5** **Self Tutor**

The table below shows the numbers of aces served by tennis players in their first set of a tournament.

| *Number of aces* | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| *Frequency* | 4 | 11 | 18 | 13 | 7 | 2 |

Determine the mean number of aces for these sets.


Serena Williams► Serves 4 ACES in a row @ Wimbledon 2012

https://www.youtube.com/watch?v=OKOWXKbXzmo

## D MEASURING THE CENTRE OF A DATA SET

**Example 5** **Self Tutor**

The table below shows the numbers of aces served by tennis players in their first set of a tournament.

| *Number of aces* | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| *Frequency* | 4 | 11 | 18 | 13 | 7 | 2 |

Determine the mean number of aces for these sets.

| *Number of aces* | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| *Frequency* | 4 | 11 | 18 | 13 | 7 | 2 | 55 |
| Product | 4 | 22 | 54 | 52 | 35 | 12 | 179 |

$$\bar{x} = \frac{\Sigma}{n} = \frac{179}{55} \approx 3.25$$

https://www.youtube.com/watch?v=OKOWXKbXzmo

## D MEASURING THE CENTRE OF A DATA SET

**Example 5** — Self Tutor

The table below shows the numbers of aces served by tennis players in their first set of a tournament.

| *Number of aces* | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| *Frequency* (# of times players had this many aces) | 4 | 11 | 18 | 13 | 7 | 2 |

Determine the mean number of aces for these sets.

| *Number of aces* | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| *Frequency* | 4 | 11 | 18 | 13 | 7 | 2 |
| Product | 4 | 22 | 54 | 52 | 35 | 12 |

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

$\bar{x} = \frac{179}{55}$

$n = 55$ players  $\Sigma = 179$

$\bar{x} \approx 3.25$ aces

| *Number of aces* | *Frequency* | *Product* |
|---|---|---|
| 1 | 4 | 4 |
| 2 | 11 | 22 |
| 3 | 18 | 54 |
| 4 | 13 | 52 |
| 5 | 7 | 35 |
| 6 | 2 | 12 |
| *Total* | 55 | 179 |

This is the number of data values. (55)

This is the sum of the data values. (179)

$$\bar{x} = \frac{\text{sum of the data values}}{\text{number of data values}}$$
$$= \frac{179}{55}$$
$$\approx 3.25 \text{ aces}$$

## D MEASURING THE CENTRE OF A DATA SET

What if the data is grouped, (we do not know the exact data values)??

**Example 9** Self Tuto

The speeds of 129 cars were measured at a particular point on a country road. They are shown in the table alongside.

Estimate the mean speed of the drivers.

| *Speed* $v$ (km/h) | *Frequency* |
|---|---|
| $40 \leqslant v < 50$ | 1 |
| $50 \leqslant v < 60$ | 3 |
| $60 \leqslant v < 70$ | 17 |
| $70 \leqslant v < 80$ | 39 |
| $80 \leqslant v < 90$ | 48 |
| $90 \leqslant v < 100$ | 17 |
| $100 \leqslant v < 110$ | 4 |

# What if the data is grouped, (we do not know the exact data values)??

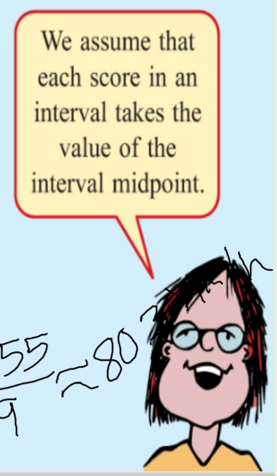
| *Speed v* (km/h) | *Frequency* | *Interval midpoint* | *Product* |
|---|---|---|---|
| $40 \leqslant v < 50$ | 1 | 45 | 45 |
| $50 \leqslant v < 60$ | 3 | 45 | 45 |
| $60 \leqslant v < 70$ | 17 | 55 | 165 |
| $70 \leqslant v < 80$ | 39 | 65 | 1105 |
| $80 \leqslant v < 90$ | 48 | 75 | 2925 |
| $90 \leqslant v < 100$ | 17 | 85 | 4080 |
| $100 \leqslant v < 110$ | 4 | 95 | 1615 |
| *Total* | 129 | 105 | 420 |

10,355

$$\therefore \quad \text{mean} = \frac{\text{sum of the data values}}{\text{number of data values}}$$

$$\approx \frac{10\,355}{129}$$

$$\bar{x} \approx \frac{10{,}355}{129} \approx 80$$



## D MEASURING THE CENTRE OF A DATA SET

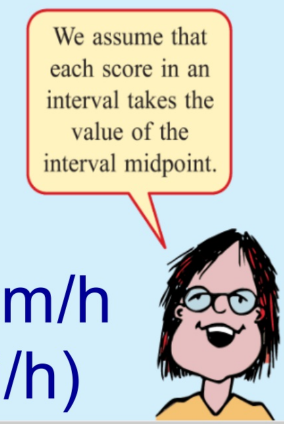
What if the data is grouped, (we do not know the exact data values)??

| *Speed* $v$ (km/h) | *Frequency* | *Interval midpoint* | *Product* |
|---|---|---|---|
| $40 \leqslant v < 50$ | 1 | 45 | 45 |
| $50 \leqslant v < 60$ | 3 | 55 | 165 |
| $60 \leqslant v < 70$ | 17 | 65 | 1105 |
| $70 \leqslant v < 80$ | 39 | 75 | 2925 |
| $80 \leqslant v < 90$ | 48 | 85 | 4080 |
| $90 \leqslant v < 100$ | 17 | 95 | 1615 |
| $100 \leqslant v < 110$ | 4 | 105 | 420 |
| *Total* | 129 | | 10 355 |

$$\therefore \quad \text{mean} = \frac{\text{sum of the data values}}{\text{number of data values}}$$

$$\approx \frac{10\,355}{129}$$

about 80.3 km/h
(about 49 mi/h)



## Exercises

20+ minutes on two sections

**EXERCISE 10D.1**

Nongrouped data

Page 199, (3, 5, 13, 14)

**EXERCISE 10D.2**

Grouped Data

Page 203, (1, 2)

# EXERCISE 10D.1

**3** The selling prices of the last 10 houses sold in Everton Hills were:

£346 400, £327 600, £411 000, £392 500, £456 400, £332 400,
£348 000, £329 500, £331 400, £362 500

**a** Calculate the mean and median selling prices of these houses, and comment on the results.

**b** Which measure would you use if you were:

**i** a vendor wanting to sell your house **ii** looking to buy a house in the district?

**5** Packets of chocolate almonds were opened and their contents counted. The table shows the results.

| *Number in packet* | 32 | 33 | 34 | 35 | 36 | 37 | 38 |
|---|---|---|---|---|---|---|---|
| *Frequency* | 6 | 8 | 9 | 13 | 10 | 3 | 2 |

Find the mean and median of the data.

# EXERCISE 10D.1

**13** Towards the end of the season, a netballer had played 14 matches and had an average of 16.5 goals per game. In the final two matches of the season the netballer threw 21 goals and 24 goals. Find the netballer's new average.

**14** A sample of 12 measurements has a mean of 16.5, and a sample of 15 measurements has a mean of 18.6. Find the mean of all 27 measurements.

# EXERCISE 10D.2

**1** 50 adults each kicked a football. The table alongside shows the distances the ball travelled in the air before it bounced.

| *Distance* $d$ (m) | *Frequency* |
| --- | --- |
| $20 \leqslant d < 30$ | 10 |
| $30 \leqslant d < 40$ | 19 |
| $40 \leqslant d < 50$ | 14 |
| $50 \leqslant d < 60$ | 7 |

**a** Draw a frequency histogram of the data.

**b** Is it possible to determine the furthest distance the ball was kicked?

**c** Extend the table alongside to include an *interval midpoint* and *product* column.

**d** Estimate the mean distance the ball was kicked.

# EXERCISE 10D.2

**2** Sixty people were asked: "How many times have you been to the cinema in the last twelve months?". The results are given in the table alongside.

| *Number of times* | *Frequency* |
|---|---|
| 0 - 4 | 19 |
| 5 - 9 | 24 |
| 10 - 14 | 10 |
| 15 - 19 | 5 |
| 20 - 24 | 2 |

**a** Extend the table to include an *interval midpoint* and a *product* column.

**b** Estimate the mean of the data.

## EXERCISE 10D.1

**1** **a** 6 **b** 26.2 **c** $\approx 5.85$

**2** **a** 29 **b** 108 **c** 149.5

**3** **a** mean $= £363\,770$, median $= £347\,200$
The mean is affected by extreme values whereas the median is not. So the mean has been 'pushed up' higher than the median by the higher values.

**b** **i** mean selling price **ii** median selling price

**4** mean $= 5.6$ presents, median $= 6$ presents

**5** mean $\approx 34.6$ chocolate almonds,
median $= 35$ chocolate almonds

**7** **a** **A**: positively skewed; **B**: negatively skewed;
**C**: approximately symmetric

**b** **A**: mean $\approx 3.29$, median $= 3$;
**B**: mean $= 5.6$, median $= 6$;
**C**: mean $= 4.45$, median $= 4.5$

**c** **i** "For positively skewed data, the mean is *greater* than the median."

**ii** "For negatively skewed data, the mean is *less* than the median."

**iii** "For symmetric data, the mean and median are approximately *equal*."

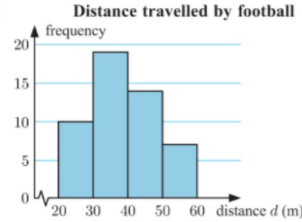
**8** 105.6 **9** 1712 km **10** \$2 592 000

**11** **a** $x = 9$ **b** $x = 12$ **12** 27 marks out of 30

**13** $\approx 17.3$ goals per match **14** $\approx 17.7$

## EXERCISE 10D.2

**1** **a** **Distance travelled by football**




**b** No; as the actual data values are lost when combining them in classes.

**c**

| *Distance d* (m) | *Frequency* | *Interval midpoint* | *Product* |
|---|---|---|---|
| $20 \leqslant d < 30$ | 10 | 25 | 250 |
| $30 \leqslant d < 40$ | 19 | 35 | 665 |
| $40 \leqslant d < 50$ | 14 | 45 | 630 |
| $50 \leqslant d < 60$ | 7 | 55 | 385 |
| *Total* | 50 | | 1930 |

**d** $\approx 38.6$ m

**2** **a**

| *Number of times* | *Frequency* | *Interval midpoint* | *Product* |
|---|---|---|---|
| 0 - 4 | 19 | 2 | 38 |
| 5 - 9 | 24 | 7 | 168 |
| 10 - 14 | 10 | 12 | 120 |
| 15 - 19 | 5 | 17 | 85 |
| 20 - 24 | 2 | 22 | 44 |
| *Total* | 60 | | 455 |

**b** $\overline{x} \approx 7.58$ times

**3** **a** $600 \leqslant A < 700$ m$^2$ **b** $\overline{x} \approx 661$ m$^2$