KING EDWARD VI
HANDSWORTH GRAMMAR
SCHOOL FOR BOYS

KING EDWARD VI
ACADEMY TRUST
BIRMINGHAM

Year 12
Applied Mathematics
S1 2 Measures of Location and Spread
Booklet

HGS Maths

Dr Frost Course

Name: ______________________________

Class: ______________________________

Contents

Experimental

i.e. Dealing with collected data.

Chp1: Data Collection

Methods of sampling, types of data, and populations vs samples.

Chp2: Measures of Location/Spread

Statistics used to summarise data, including mean, standard deviation, quartiles, percentiles. Use of linear interpolation for estimating medians/quartiles.

Chp3: Representation of Data

Producing and interpreting visual representations of data, including box plots and histograms.

Chp4: Correlation

Measuring how related two variables are, and using linear regression to predict values.

Theoretical

Deal with probabilities and modelling to make inferences about what we 'expect' to see or make predictions, often using this to reason about/contrast with experimentally collected data.

Chp5: Probability

Venn Diagrams, mutually exclusive + independent events, tree diagrams.

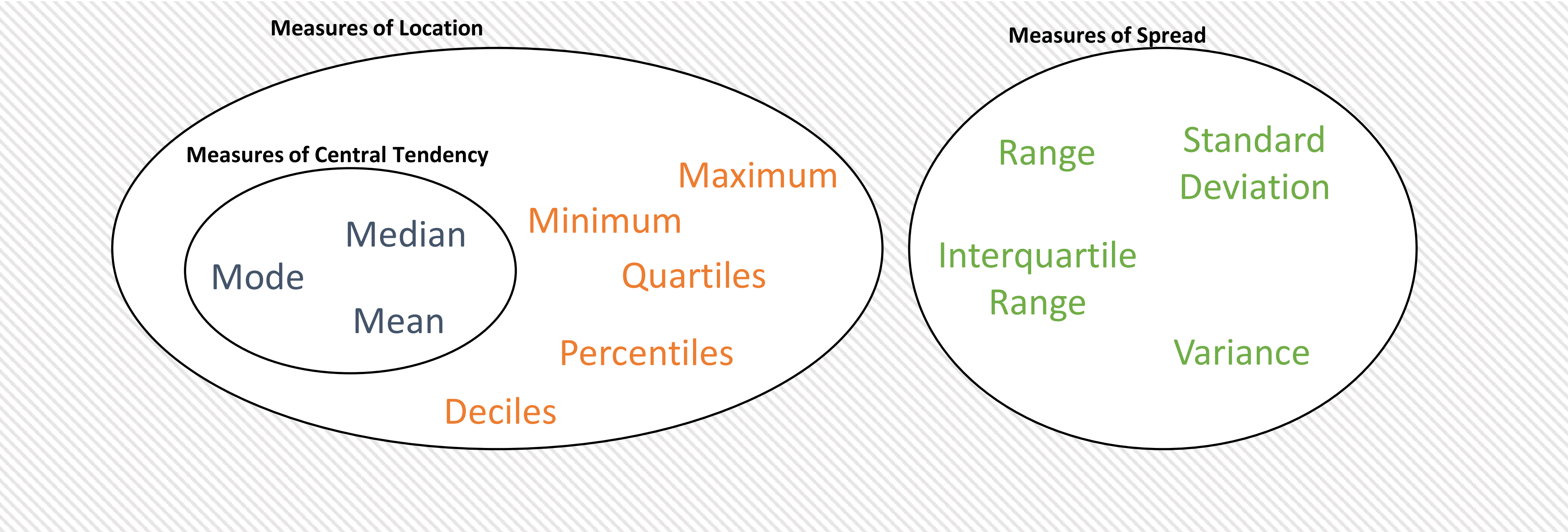
Chp6: Statistical Distributions

Common distributions used to easily find probabilities under certain modelling conditions, e.g. binomial distribution.

Chp7: Hypothesis Testing

Determining how likely observed data would have happened 'by chance', and making subsequent deductions.

2.1 Measures of Central Tendency



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**.

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 Σx 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).

Notes

Worked Example

Rebecca records the shirt collar size, x, of the male students in her year. The results are shown in the table.

Shirt collar size	15	15.5	16	16.5	17
Number of students	3	17	29	34	12

Find for this data:

a) The mode

b) The median

c) The mean

d) Explain why a shirt manufacturer might use the mode when planning production numbers.

T.22 2A: Qs all,

Worked Example

The length, x mm, to the nearest mm, of a random sample of pine cones is measured. The data is shown in the table.

Length of pine cone (mm)	30 – 31	32 – 33	34 – 36	37 – 39
Number of students	2	25	30	13

a) Write down the modal class
b) Estimate the mean
c) Find the median class

T.24 2B: Qs all, P.8 2.1: Qs all

2.2 Other Measures of Location

✐ 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.

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

Notes

Quickfire Questions...

What position do we use for the median?

Lengths: 3cm, 5cm, 6cm, …
$n = 11$

Median position:

Lengths: 4m, 8m, 12.4m, …
$n = 24$

Median position:

Age	Freq
$10 \leq a < 20$	12
$20 \leq a < 30$	5

Median position:

Score	Freq
$150 \leq s < 200$	3
$200 \leq s < 400$	7

Median position:

Ages: 5, 7, 7, 8, 9, 10, …
$n = 60$

Median position:

Score	Freq
$150 \leq s < 200$	15
$200 \leq s < 400$	6

Median position:

Weights: 1.2kg, 3.3kg, …
$n = 35$

Median position:

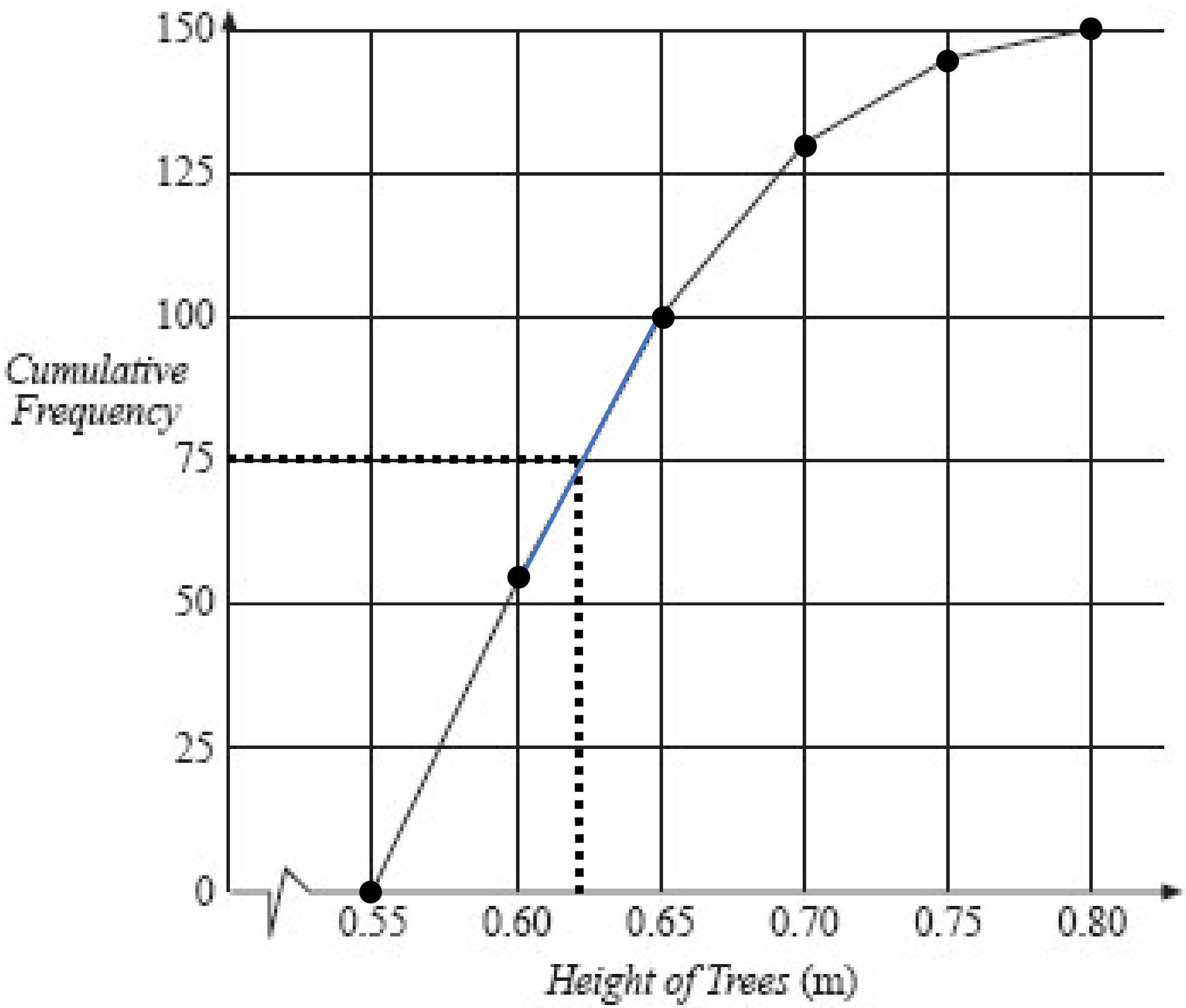
Volume (ml)	Freq
$0 \leq v < 100$	5
$100 \leq v < 200$	6
$200 \leq v < 300$	2

Median position:

Weights: 4.4kg, 7.6kg, 7.7kg…
$n = 18$

Median position:

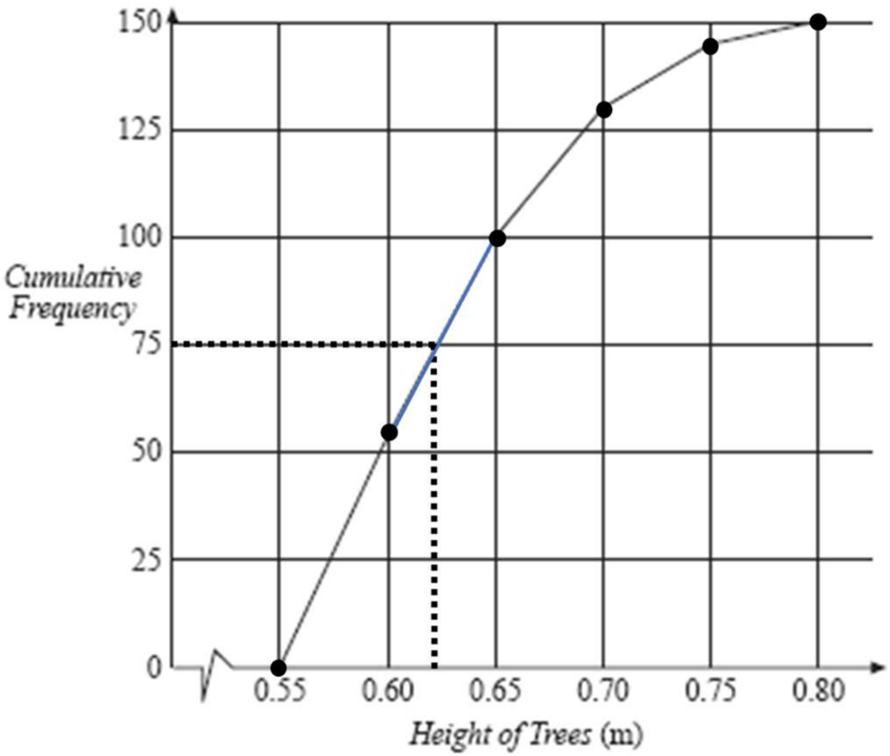
Linear Interpolation



Height of tree (m)	Freq	C.F.
$0.55 \leq h < 0.6$	55	55
$0.6 \leq h < 0.65$	45	100
$0.65 \leq h < 0.7$	30	130
$0.7 \leq h < 0.75$	15	145
$0.75 \leq h < 0.8$	5	150

At GCSE we could find the median by drawing a suitable line on a cumulative frequency graph. How could we read off this value exactly using a suitable calculation?

Linear Interpolation



Height of tree (m)	Freq	C.F.
$0.55 \leq h < 0.6$	55	55
$0.6 \leq h < 0.65$	45	100
$0.65 \leq h < 0.7$	30	130
$0.7 \leq h < 0.75$	15	145
$0.75 \leq h < 0.8$	5	150

The 75th item is within the $0.6 \leq h < 0.65$ class interval because 75 is within the first 100 items but not the first 55.

Frequency up until this interval

Item number we're interested in.

Frequency by end of this interval

Med

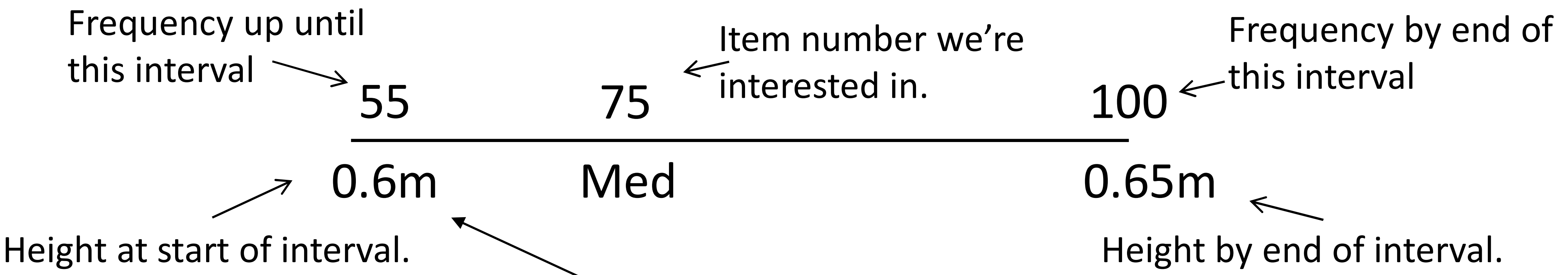
Height at start of interval.

Height by end of interval.

Linear Interpolation

Height of tree (m)	Freq	C.F.
$0.55 \leq h < 0.6$	55	55
$0.6 \leq h < 0.65$	45	100
$0.65 \leq h < 0.7$	30	130
$0.7 \leq h < 0.75$	15	145
$0.75 \leq h < 0.8$	5	150

What fraction of the way across the interval are we?



Fro Tip: I like to put the units to avoid getting frequencies confused with values of the variable.

Fro Tip: To quickly get frequency before and after, just look for the two cumulative frequencies that surround the item number.

More Examples Finding the median

Weight of cat (kg)	Freq	C.F.
$1.5 \leq w < 3$	10	10
$3 \leq w < 4$	8	18
$4 \leq w < 6$	14	32

Time (s)	Freq	C.F.
$8 \leq t < 10$	4	4
$10 \leq t < 12$	3	7
$12 \leq t < 14$	13	20

Worked Example

544d: **Estimate the median of grouped data using interpolation.**

Jenny collects the heights of 60 flowers and records the data in the table below.

Height (z cm)	**Frequency**
$70 < z \leq 80$	3
$80 < z \leq 90$	6
$90 < z \leq 100$	20
$100 < z \leq 110$	10
$110 < z \leq 120$	8
$120 < z \leq 130$	13

Use interpolation to estimate the median.
Give your answer correct to 1 decimal place.

544f: **Determine quartiles or percentiles of grouped data using interpolation.**

John collects the heights of 90 flowers and records the data in the table below.

Height (z cm)	Frequency
$15 < z \leq 20$	5
$20 < z \leq 25$	13
$25 < z \leq 30$	55
$30 < z \leq 35$	10
$35 < z \leq 40$	7

Use interpolation to estimate the upper quartile.
Give your answer correct to 1 decimal place.

What's different about the intervals here?

Weight of cat to nearest kg	Frequency
$10 - 12$	7
$13 - 15$	2
$16 - 18$	9
$19 - 20$	4

There are **GAPS** between intervals!
What interval does this **actually** represent?

$$10 - 12$$

$$9.5 - 12.5$$

Lower class boundary

Upper class boundary

Class width = 3

Identify the class width

Distance d travelled (in m)	…
$0 \leq d < 150$	
$150 \leq d < 200$	
$\mathbf{200 \leq d < 210}$	

Lower class boundary =

Class width =

Time t taken (in seconds)	…
$0 - 3$	
$\mathbf{4 - 6}$	
$7 - 11$	

Lower class boundary =

Class width =

Weight w in kg	…
$10 - 20$	
$21 - 30$	
$\mathbf{31 - 40}$	

Lower class boundary =

Class width =

Speed s (in mph)	…
$10 \leq s < 20$	
$20 \leq s < 29$	
$\mathbf{29 \leq s < 31}$	

Lower class boundary =

Class width =

Linear Interpolation with gaps

Edexcel S1 Jan 2007 Q4

Summarised below are the distances, to the nearest mile, travelled to work by a random sample of 120 commuters.

Distance (to the nearest mile)	Number of commuters	
0 – 9	10	10
10 – 19	19	29
20 – 29	43	72
30 – 39	25	97
40 – 49	8	105
50 – 59	6	111
60 – 69	5	116
70 – 79	3	119
80 – 89	1	120

For this distribution,

~~(*a*) describe its shape, **(1)**~~

(*b*) use linear interpolation to estimate its median. **(2)**

Test Your Understanding – Find the median

Age of relic (years)	Frequency
0-1000	24
1001-1500	29
1501-1700	12
1701-2000	35

Shark length (cm)	Frequency
$40 \leq x < 100$	17
$100 \leq x < 300$	5
$300 \leq x < 600$	8
$600 \leq x < 1000$	10

Worked Example

544e: **Estimate the median of grouped data where intervals have gaps.**

John collects the lengths of 90 animals and records the data in the table below.

Length (x cm)	Frequency
$50 - 59$	8
$60 - 69$	14
$70 - 79$	26
$80 - 89$	13
$90 - 99$	9
$100 - 109$	20

Use interpolation to estimate the median.
Give your answer correct to 1 decimal place.

T.27 2C: Qs 2-4,7 , P.9 2.2: Qs all

Supplementary Exercise 1

1

The number of patients attending a hospital trauma clinic each day was recorded over several months, giving the data in the table below.

Number of patients	10 - 19	20 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 69
Frequency	2	18	24	30	27	14	5

Use linear interpolation to estimate the median of these data.

2

The ages of 300 houses in a village are recorded given the following table of results.

Age a (years)	Number of houses
$0 \le a < 20$	36
$20 \le a < 40$	92
$40 \le a < 60$	74
$60 \le a < 100$	39
$100 \le a < 200$	14
$200 \le a < 300$	27
$300 \le a < 500$	18

Use linear interpolation to estimate the median.

3

A cyber-café recorded how long each user stayed during one day giving the following results.

Length of stay (minutes)	Number of houses
$0 \le l < 30$	15
$30 \le l < 60$	31
$60 \le l < 90$	32
$90 \le l < 120$	23
$120 \le l < 240$	17
$240 \le l < 360$	2

Use linear interpolation to estimate the median of these data.

4

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(*a*) Estimate the mean ~~and standard deviation~~ of these data. **(5)**

(*b*) Use linear interpolation to estimate the value of the median. **(2)**

Supplementary Exercise 1

1

Questions should be on a printed sheet…

The number of patients attending a hospital trauma clinic each day was recorded over several months, giving the data in the table below.

Number of patients	10 - 19	20 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 69
Frequency	2	18	24	30	27	14	5

Use linear interpolation to estimate the median of these data.

$$\boldsymbol{Median = 34.5 + \left(\frac{16}{30} \times 5\right) = 37.2}$$

2

The ages of 300 houses in a village are recorded given the following table of results.

Age a (years)	Number of houses
$0 \leq a < 20$	36
$20 \leq a < 40$	92
$40 \leq a < 60$	74
$60 \leq a < 100$	39
$100 \leq a < 200$	14
$200 \leq a < 300$	27
$300 \leq a < 500$	18

Use linear interpolation to estimate the median.

$$\boldsymbol{Median = 40 + \left(\frac{22}{74} \times 20\right) = 45.9}$$

3

A cyber-café recorded how long each user stayed during one day giving the following results.

Length of stay (minutes)	Number of houses
$0 \leq l < 30$	15
$30 \leq l < 60$	31
$60 \leq l < 90$	32
$90 \leq l < 120$	23
$120 \leq l < 240$	17
$240 \leq l < 360$	2

Use linear interpolation to estimate the median of these data.

$$\boldsymbol{Median = 60 + \left(\frac{14}{32} \times 30\right) = 73.125}$$

2.3 Measures of Spread

✎ To find the position of the LQ/UQ for listed data, find $\frac{1}{4}n$ or $\frac{3}{4}n$ then as before:

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

✎ To find the LQ and UQ of grouped data, find $\frac{1}{4}n$ and $\frac{3}{4}n$, then use linear interpolation.

Notes

Percentiles

The LQ, median and UQ give you 25%, 50% and 75% along the data respectively.
But we can have any percentage you like!

$n = 43$
Item to use for 57th percentile? ➡

You will always find these for grouped data in an exam, **so never round this position**.

Notation:

Lower Quartile:

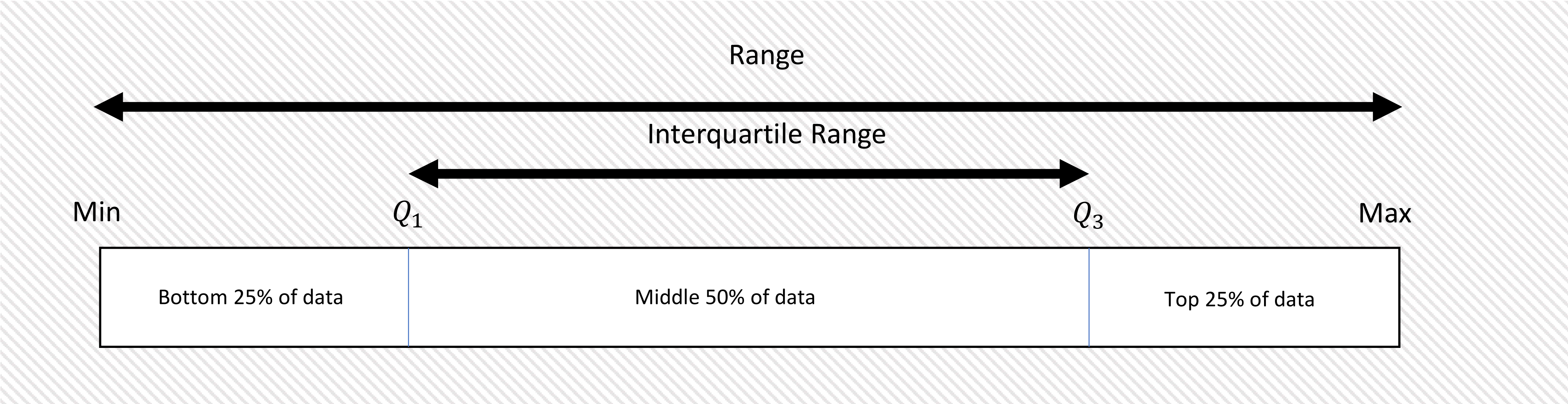
Median:

Upper Quartile:

57th Percentile:

Measures of Spread

The interquartile range and interpercentile range are examples of **measures of spread.**



Interquartile Range = Upper Quartile − Lower Quartile

Why might we favour the interquartile range over the range?
Because it gives us the spread of the data excluding the extreme values at either end.

We can control this further by having for example the "*10th to 90th interpercentile range*", which would be $P_{90} - P_{10}$. This would typically be symmetrical about the median, so that we could interpret this as "*the range of the data with the most extreme 10% of values at either end excluded*".
The 10th percentile is also known as the 1st **decile** (D_1), and similarly $P_{90} = D_9$.

Exercise

Age of relic (years)	Frequency
0-1000	24
1001-1500	29
1501-1700	12
1701-2000	35

Item for Q_1:

$Q_1 =$

$Q_3 =$

$IQR =$

Shark length (cm)	Frequency
$40 \leq x < 100$	17
$100 \leq x < 300$	5
$300 \leq x < 600$	8
$600 \leq x < 1000$	11

Item to use for P_{10}:

$P_{10} =$

$P_{90} =$

10^{th} to 90^{th} interpercentile range:

Worked Example

The table shows the masses, in tonnes, of 120 African bush elephants.

Mass, m (t)	$4.0 \leq m < 4.5$	$4.5 \leq m < 5.0$	$5.0 \leq m < 5.5$	$5.5 \leq m < 6.0$	$6.0 \leq m < 6.5$
Frequency	13	23	31	34	19

Find estimate for:

a) The range

b) The interquartile range

Worked Example

The table shows the masses, in tonnes, of 120 African bush elephants.

Mass, m (t)	$4.0 \leq m < 4.5$	$4.5 \leq m < 5.0$	$5.0 \leq m < 5.5$	$5.5 \leq m < 6.0$	$6.0 \leq m < 6.5$
Frequency	13	23	31	34	19

Find estimate for the 10^{th} to 90^{th} interpercentile range.

T.29 2D: Qs all , P.10 2.3: all

Supplementary Exercise 2

Q1) May 2013 Q4 (continued)

The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

(*c*) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures. **(1)**

(*d*) Estimate the interquartile range of this distribution. **(2)**

Q3)

The ages of 300 houses in a village are recorded given the following table of results.

Age a (years)	Number of houses
$0 \le a < 20$	36
$20 \le a < 40$	92
$40 \le a < 60$	74
$60 \le a < 100$	39
$100 \le a < 200$	14
$200 \le a < 300$	27
$300 \le a < 500$	18

Use linear interpolation to estimate the lower quartile, upper quartile and hence the interquartile range.

Q4)

A cyber-café recorded how long each user stayed during one day giving the following results.

Length of stay (minutes)	Number of houses
$0 \le l < 30$	15
$30 \le l < 60$	31
$60 \le l < 90$	32
$90 \le l < 120$	23
$120 \le l < 240$	17
$240 \le l < 360$	2

Use linear interpolation to estimate:

a) The lower quartile.

b) The upper quartile.

c) The 90th percentile.

Q2) June 2005 Q2

The following table summarises the distances, to the nearest km, that 134 examiners travelled to attend a meeting in London.

Distance (km)	Number of examiners
41–45	4
46–50	19
51–60	53
61–70	37
71–90	15
91–150	6

(*c*) Use interpolation to estimate the median Q_2, the lower quartile Q_1, and the upper quartile Q_3 of these data.

Q5)

Distance (to the nearest mile)	Number of commuters
0 – 9	10
10 – 19	19
20 – 29	43
30 – 39	25
40 – 49	8
50 – 59	6
60 – 69	5
70 – 79	3
80 – 89	1

Find the interquartile range for the distance travelled by commuters.

2.4 Variance and Standard Deviation

Variance is a measure of spread that takes all values into account. Variance, by definition, is the **average squared distance from the mean**.

Distance from mean…

Squared distance from mean…

Average squared distance from mean…

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

Notation: While Σ is 'uppercase sigma' and means 'sum of', σ is 'lowercase sigma' (we'll see why we have the squared in a sec)

But in practice you will never use this form, and it's possible to simplify the formula to the following:*

"The mean of the squares minus the square of the mean ('msmsm')"

$$\sigma^2 = \frac{\Sigma x^2}{n} - \bar{x}^2$$

Standard Deviation: $\sigma = \sqrt{Variance}$

The standard deviation **can 'roughly' be thought of as the <u>average distance from the mean</u>.**

Extending to frequency/grouped frequency tables:

$$Variance = \frac{\Sigma f x^2}{\Sigma f} - \bar{x}^2$$

It's better to try and memorise the mnemonic than the formula itself – you'll understand what's going on better. In an exam, you will pretty much certainly be asked to find the standard deviation for grouped data, and not listed data.

Notes

Worked Example

534c: **Calculate the standard deviation of listed data, using formulae.**

A data set has the following 6 items:

31 26 39 30 33 21

Calculate the variance.

$$\text{Standard deviation} = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

Variance =

Worked Example

534f: **Calculate the variance or standard deviation given $\sum x$ and $\sum x^2$**

A data set with 21 items has the following statistics:

$\sum x = 596$ and $\sum x^2 = 17536$

Calculate the standard deviation.

Standard deviation = √(Variance)

Standard deviation = $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$

Worked Example

534g: Determine the variance or standard deviation of grouped data.

Jenny collects the running times of some athletes and records the data in the table below.

Time (z seconds)	Frequency
$70 < z \leq 80$	1
$80 < z \leq 90$	7
$90 < z \leq 100$	6
$100 < z \leq 110$	2
$110 < z \leq 120$	1

Find the standard deviation of the data in the table. Give your answer correct to 1 decimal place where appropriate.

Worked Example

The scores, x, were recorded for 20 people.
The summary data is $\sum x = 34$, $\sum x^2 = 567$
The highest score was 8.5.
The lowest score was 0.2.
Estimate the number of scores which were greater than one standard deviation above the mean.

Most common exam errors

- Thinking $\Sigma f x^2$ means $(\Sigma f x)^2$. It means the sum of each value squared!
- When asked to calculate the mean followed by standard deviation, using a rounded version of the mean in calculating the standard deviation, and hence introducing rounding errors.
- Forgetting to square root the variance to get the standard deviation.

ALL these mistakes can be easily spotted if you check your value against "σx" in STATS mode.

T.32 2E all Qs, P.12 2.4: all Qs

2.5 Coding

Suppose our original variable (e.g. heights in cm) was x. Then y would represent the heights with 10cm added on to each value.

Coding	Effect on $\bar{x}$	Effect on σ
$y = x + 10$	$\bar{x}$ will similarly increase by 10 (to get $\bar{y}$)	As discussed, adding (and subtracting) has no effect on standard deviation or any measure of spread.
$y = 3x$	$\bar{x}$ will get 3 times bigger.	Standard deviation will get 3 times larger.
$y = 2x - 5$	$\bar{y} = 2\bar{x} - 5$, i.e. effect on values is same effect on mean.	-5 has no effect but standard deviation will get 2 times larger.

You might get any **linear** coding (i.e. using $\times + \div -$). We might think that any operation on the values has the same effect on the mean. But note for example that **squaring** the values would not square the mean; we already know that $\Sigma x^2 \neq (\Sigma x)^2$ in general.

Notes

Quickfire Questions

Old mean $\bar{x}$	Old σ_x	Coding	New mean $\bar{y}$	New σ_y
36	4	$y = x - 20$		
		$y = 2x$	72	16
35	4	$y = 3x - 20$		
		$y = \frac{x}{2}$	20	$\frac{3}{2}$
11	27	$y = \frac{x + 10}{3}$		
		$y = \frac{x - 100}{5}$	40	5

Worked Example

545c: **Calculate the mean of coded data with a single transformation.**

The daily mean pressure x in Leeming is recorded over a selected period. The mean is 997.1 and the standard deviation is 70.8.

The data is coded using $y = \frac{4x-5}{9}$

Find the mean and standard deviation of the coded data. Give your answers correct to 3 significant figures.

$\bar{y} =$ ______ hPa

$\sigma_y =$ ______ hPa

Worked Example

545f: Determine the mean and standard deviation given the coded mean and standard deviation.

The daily mean temperature x in Jacksonville is recorded over a selected period.

The data is coded using $y = \frac{2x-47}{5}$

The mean of the coded data is 2.08 and the standard deviation of the coded data is 0.960.

Find the mean and standard deviation of x.
Give your answers correct to 3 significant figures.

$\bar{x} =$ ____ °

$\sigma_x =$ ____ °

Worked Example

A scientist measures the temperature, x °C, at five different points in a nuclear reactor. Her results are given below:
332°C, 355°C, 306°C, 317°C, 340°C

a) Use the coding $y = \frac{x-300}{10}$ to code this data.

b) Calculate the mean and standard deviation of the coded data.

c) Use your answer to part b to calculate the mean and standard deviation of the original data.

Worked Example

From the large data set, data on the maximum gust, g knots, is recorded in Leuchars during May and June 2015.
The data was coded using $h = \frac{g-5}{1o}$ and the following statistics found:

$S_{hh} = 43.58$

$\bar{h} = 2$

$n = 61$

Calculate the mean and standard deviation of the maximum gust in knots.

T.34 2F all Qs, P.13 2.5: all Qs

Standard deviation

Standard deviation = √(Variance)

Interquartile range = $IQR = Q_3 - Q_1$

For a set of n values $x_1, x_2, \ldots x_i, \ldots x_n$

$$S_{xx} = \Sigma(x_i - \bar{x})^2 = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$$

Standard deviation = $\sqrt{\frac{S_{xx}}{n}}$ or $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$

3. Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, x, using the formula $y = x - 1010$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \qquad \sum y^2 = 5912$$

(a) State the units of the variable x **(1)**

(b) Find the mean Daily Mean Pressure for these 30 days. **(2)**

(c) Find the standard deviation of Daily Mean Pressure for these 30 days. **(3)**

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

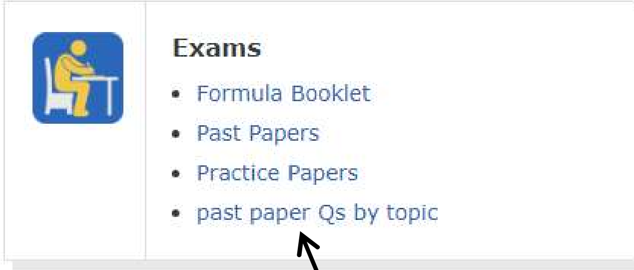
Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1029	1028	1028
Cardinal Wind Direction			

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

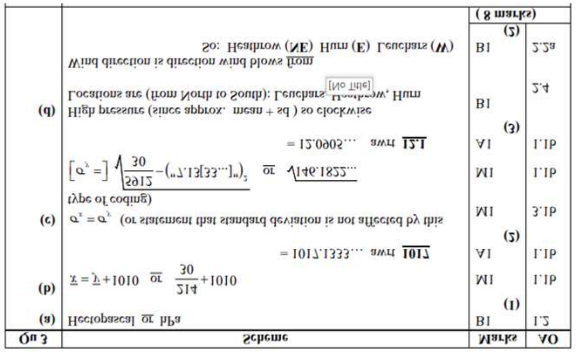
W NE E

You may assume that these 3 locations were under a single region of pressure.

(d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table.
Give a reason for your answer. **(2)**



Past paper practice by topic. Both new and old specification can be found via this link on hgsmaths.com



Qu 3	Scheme	Marks	AO
(a)	Hectopascal or hPa	B1 (1)	1.2
(b)	$\bar{x} = \bar{y} + 1010$ or $\frac{214}{30} + 1010$	M1	1.1b
	$= 1017.1333\ldots$ awrt **1017**	A1 (2)	1.1b
(c)	$\sigma_x = \sigma_y$ (or statement that standard deviation is not affected by this type of coding)	M1	3.1b
	$[\sigma_y =]\ \sqrt{\frac{5912}{30} - (\text{"}7.13[33\ldots]\text{"})^2}$ or $\sqrt{146.1822\ldots}$	M1	1.1b
	$= 12.0905\ldots$ awrt **12.1**	A1 (3)	1.1b
(d)	High pressure (since approx. mean + sd) so clockwise Locations are (from North to South): Leuchars, Heathrow, Hurn	B1	2.4
	Wind direction is direction wind blows from So: Heathrow (**NE**) Hurn (**E**) Leuchars (**W**)	B1 (2)	2.2a
		(8 marks)	

Summary of Key Points

1 The **mode** or **modal class** is the value or class that occurs most often.

2 The **median** is the middle value when the data values are put in order.

3 The **mean** can be calculated using the formula $\bar{x} = \frac{\Sigma x}{n}$.

4 For data given in a frequency table, the mean can be calculated using the formula $\bar{x} = \frac{\Sigma xf}{\Sigma f}$.

5 To find the **lower quartile** for discrete data, divide n by 4. If this is a whole number, the lower quartile is halfway between this data point and the one above. If it is not a whole number, round *up* and pick this data point.

6 To find the **upper quartile** for discrete data, find $\frac{3}{4}$ of n. If this is a whole number, the upper quartile is halfway between this data point and the one above. If it is not a whole number, round *up* and pick this data point.

7 The **range** is the difference between the largest and smallest values in the data set.

8 The **interquartile range** (IQR) is the difference between the upper quartile and the lower quartile, $Q_3 - Q_1$.

9 The **interpercentile range** is the difference between the values for two given percentiles.

10 **Variance** $= \frac{\Sigma(x-\bar{x})^2}{n} = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2 = \frac{S_{xx}}{n}$ where $S_{xx} = \Sigma(x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$

11 The **standard deviation** is the square root of the variance:

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2} = \sqrt{\frac{S_{xx}}{n}}$$

12 You can use these versions of the formulae for variance and standard deviation for grouped data that is presented in a frequency table:

$$\sigma^2 = \frac{\Sigma f(x-\bar{x})^2}{\Sigma f} = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2 \qquad \sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$$

where f is the frequency for each group and Σf is the total frequency.

13 If data is coded using the formula $y = \frac{x-a}{b}$

- the mean of the coded data is given by $\bar{y} = \frac{\bar{x}-a}{b}$
- the standard deviation of the coded data is given by $\sigma_y = \frac{\sigma_x}{b}$ where σ_x is the standard deviation of the original data.

T.36 mixed ex., P.14 BSG