Stats1 Chapter 2 :: Measures of Location & Spread

jfrost@tiffin.kingston.sch.uk

www.drfrostmaths.com
@DrFrostMaths

Last modified: 7th June 2018

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.

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 where 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.

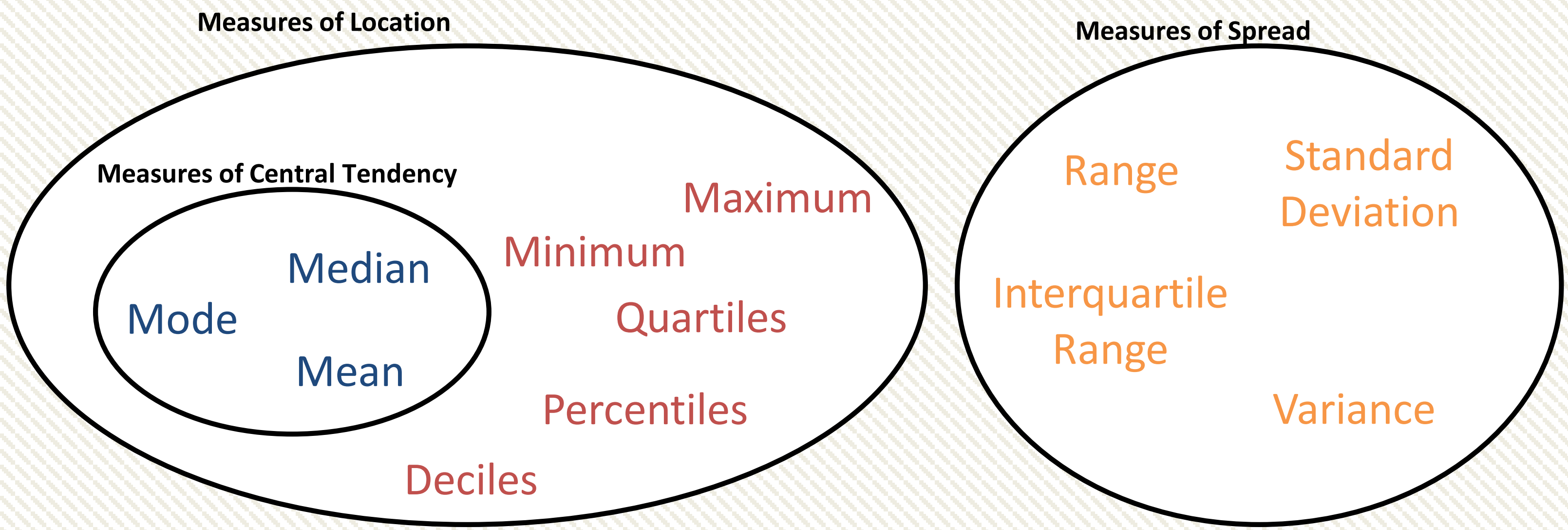
Differences

- 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 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).

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.

$$\bar{x} = \frac{?}{?}$$

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.

$n =$?

$\Sigma x / n =$?

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

$\Sigma x =$?

$\bar{x} =$?

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: $\bar{x} =$? $=$? $=$?

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 "$0 \times 4 + \cdots$"

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…

$\Sigma f x$ → ?

Σf → ?

Opinion: <u>I wouldn't even bother remembering the $\bar{x} = \frac{\Sigma f x}{\Sigma f}$ formula</u>.
I'd instead just remember $\bar{x} = \frac{\Sigma x}{n}$ and just think what the Σx 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: 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

$\bar{x} =$?

Exercises 2A/2B

Pearson Statistics & Mechanics Year 1/AS
Pages 22-23, 24-25

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.

a) What is the overall mean of all the pupils' marks.
b) 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.

?

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?

?

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 30th. 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		
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:

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: ?

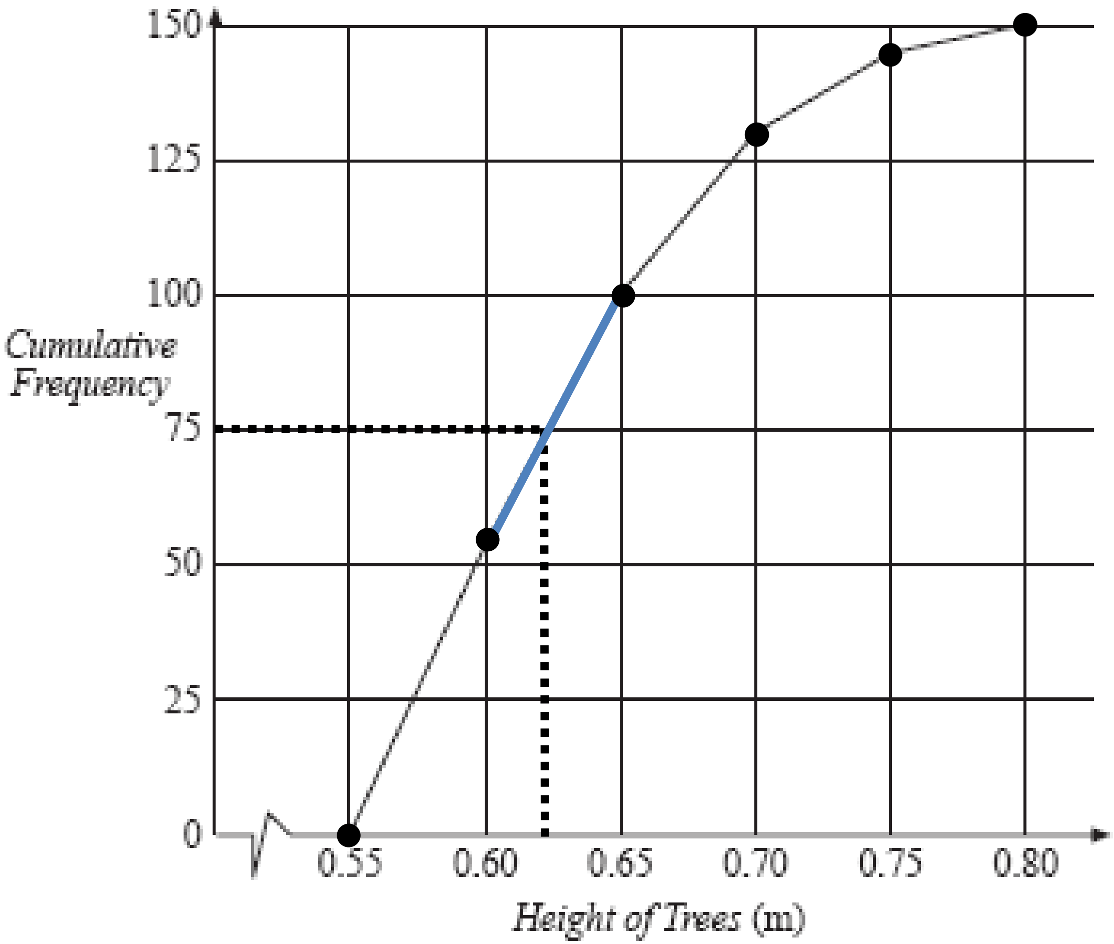
Volume (ml)	Freq
$0 \leq v < 100$	5
$100 \leq v < 200$	6
$200 \leq v < 300$	2

Median position: ?

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

Median position: ?

Linear Interpolation

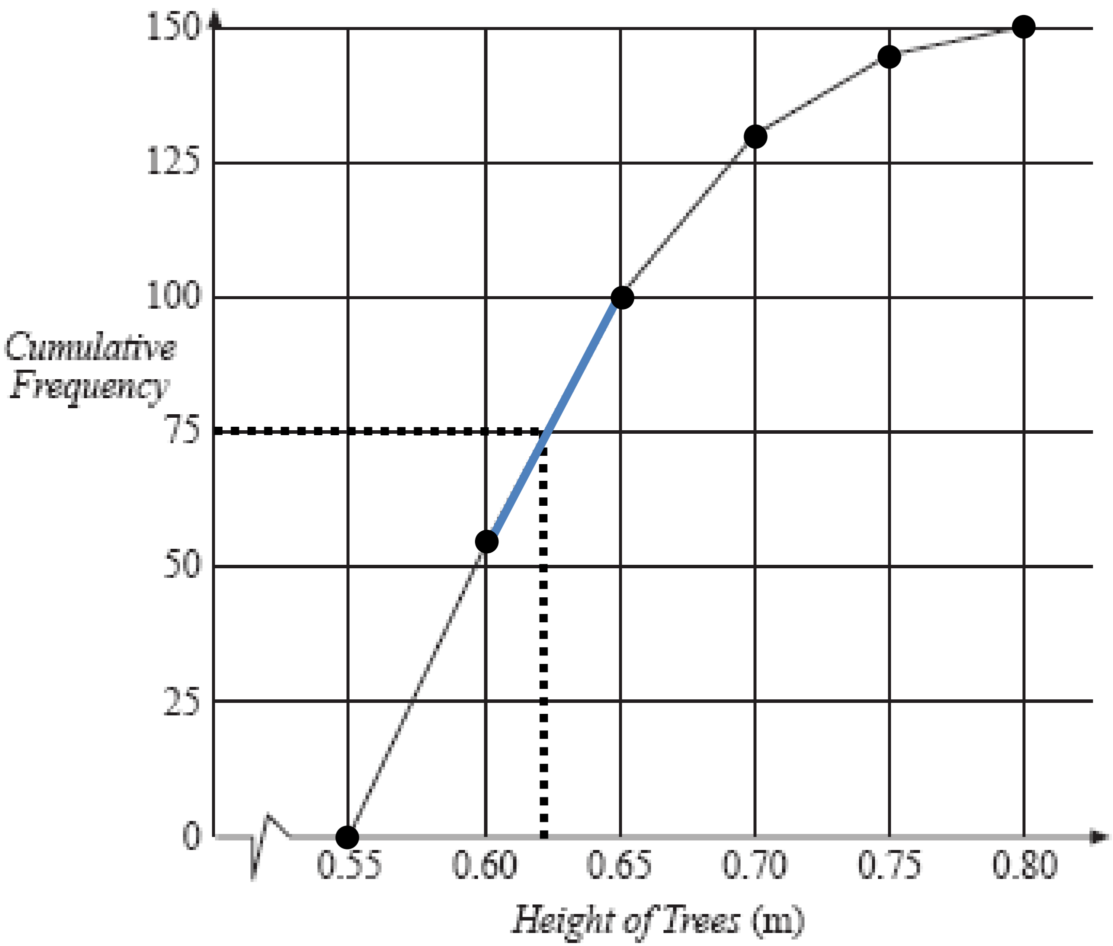


Height of tree (m)	Freq	C.F.
$0.55 \leq h < 0.6$	55	55
$0.6 \leq h < 0.65$	45	100
$0.65 \leq h < 0.7$	30	130
$0.7 \leq h < 0.75$	15	145
$0.75 \leq h < 0.8$	5	150

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

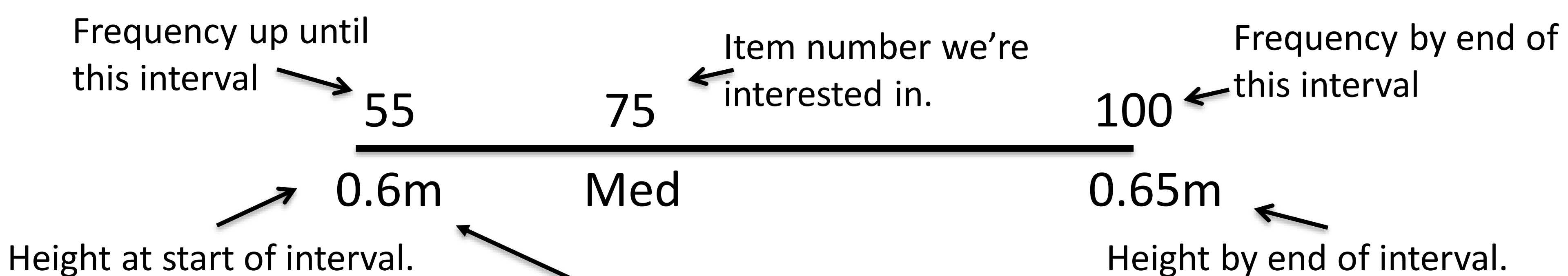
Height of tree (m)	Freq	C.F.
$0.55 \leq h < 0.6$	55	55
$0.6 \leq h < 0.65$	45	100
$0.65 \leq h < 0.7$	30	130
$0.7 \leq h < 0.75$	15	145
$0.75 \leq h < 0.8$	5	150

What fraction of the way across the interval are we?

?

Hence:

$\boldsymbol{Median} =$?



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

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

More Examples

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

Median class interval:

?

? ? ?

? Med ?

Fraction along interval:

?

Median:

?

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

? ? ?

? Med ?

Median:

?

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 **<u>GAPS</u>** between intervals!
What interval does this **actually** represent?

$$10 - 12$$

?

Lower class boundary

Upper class boundary

Class width = ?

Identify the class width

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

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)	…
$10 \leq \mathrm{s} < 20$	
$20 \leq s < 29$	
$\mathbf{29 \leq s < 31}$	

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)**

? ? ?

? Med ?

$Median = $?

Test Your Understanding

Questions should be on a printed sheet…

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

$Median =$

$=$

?

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

$Median =$

$=$

?

Exercises 2C

Pearson Statistics & Mechanics Year 1/AS
Pages 27-28

Q1a, 3, 4a, 5a

There is also a supplementary worksheet consisting of S1 exam questions (see next slides).

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.

$\boldsymbol{Median} =$?

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.

$\boldsymbol{Median} =$?

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.

$\boldsymbol{Median} =$?

Supplementary Exercise 1

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)**

$$\boldsymbol{Mean} =$$?

$$\boldsymbol{Median} =$$?

Quartiles – which item?

You need to be able to find the **quartiles** of both listed data and of grouped data. The rule is exactly the same as for the median.

Listed data

Items	n	Position of LQ & UQ	LQ & UQ
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 LQ/UQ 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 LQ:

?

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

Again, **DO NOT** round this value.

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: ?

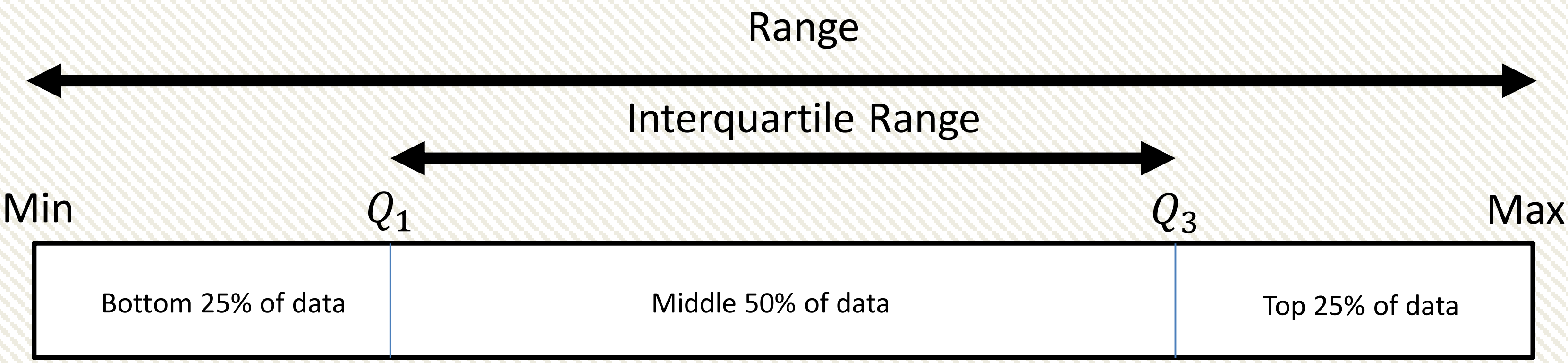
Median: ?

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?

?

Test Your Understanding

These are the same as the 'Test Your Understanding' questions on your supplementary sheet from before.

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}: ?

$P_{10} =$?

$P_{90} =$?

10th to 90th interpercentile range:

?

Exercises 2C/2D

Pearson Statistics & Mechanics Year 1/AS
Pages 27-28

Ex 2C: Q1b, 2, 4b-d, 5b-c, 6
Ex 2D

Again, there is also a supplementary worksheet consisting of S1 exam questions (see next slides).

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)**

?

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.

?

Supplementary Exercise 2

Q3)
The ages of 300 houses in a village are recorded given the following table of results.

Age a (years)	Number of houses
$0 \leq a < 20$	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 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
$0 \leq l < 30$	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:

a) The lower quartile.
b) The upper quartile.
c) The 90th percentile.

?

Q5)

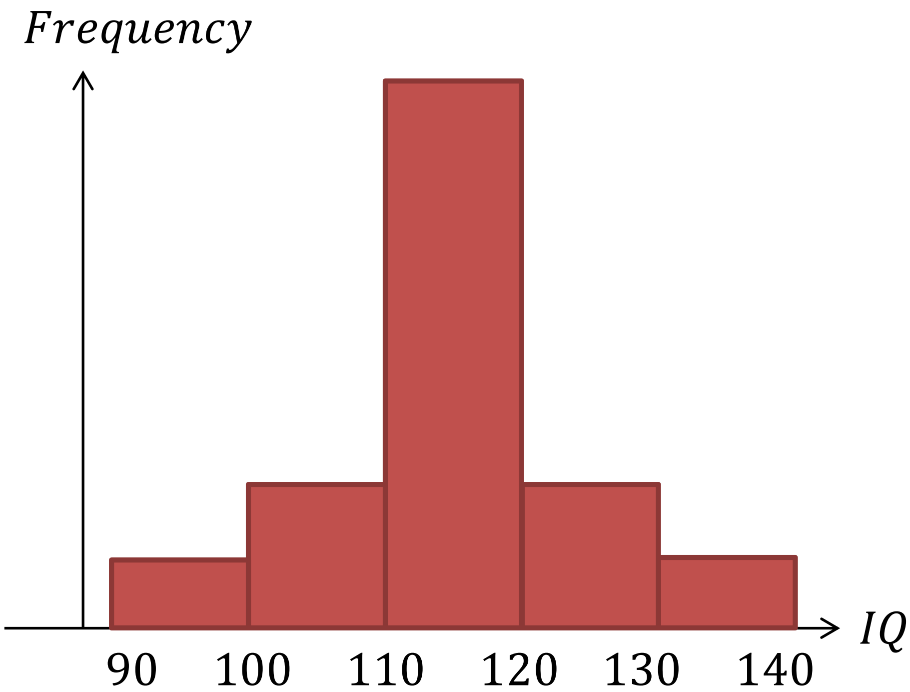
Distance (to the nearest mile)	Number of commuters
0 – 9	10
10 – 19	19
20 – 29	43
30 – 39	25
40 – 49	8
50 – 59	6
60 – 69	5
70 – 79	3
80 – 89	1

Find the interquartile range for the distance travelled by commuters.

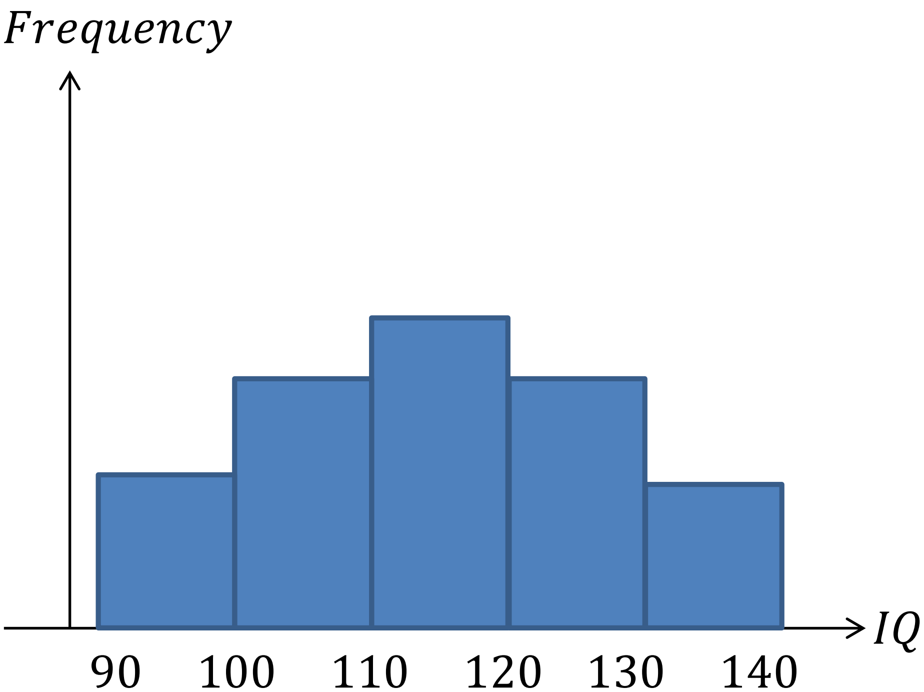
?

What is variance?

Distribution of IQs in L6Ms4



Distribution of IQs in L6Ms5



Here are the distribution of IQs in two classes of the same size. What's the same, and what's different?

?

Variance

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

Distance from mean...

Squared distance from mean...

Average squared distance from mean...

Simpler formula for variance

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

Variance

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

$$\sigma^2 = \boxed{?} - \boxed{?}$$

*** Proof: (certainly not in syllabus!)**

?

Standard Deviation

$$\sigma = \sqrt{Variance}$$

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

Examples

3, 11

Variance

$\sigma^2 =$?

Standard Deviation

$\sigma =$?

So note that that in the case of two items, the standard deviation **<u>is</u>** indeed the average distance of the values from the mean.

2cm 3cm 3cm 5cm 7cm

Variance

$\sigma^2 =$?

Standard Deviation

$\sigma =$?

Practice

Find the variance and standard deviation of the following sets of data.

2 4 6

Variance = ? Standard Deviation = ?

1 2 3 4 5

Variance = ? Standard Deviation = ?

Extending to frequency/grouped frequency tables

We can just mull over our mnemonic again:

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

$$Variance = \boxed{?} - \boxed{?}$$

Tip: It's better to try and memorise the mnemonic than the formula itself – you'll understand what's going on better.

Exam Note: In an exam, you will pretty much certainly be asked to find the standard deviation for grouped data, and not listed data.

Example

May 2013 Q4

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)**

We can use our STATS mode to work out the various summations needed (and "1-Variable Calc" will contain this amongst its list). Just input the table as normal. Note that, as per the discussion before, Σx^2 on a calculator actually gives you $\Sigma f x^2$ because it's **already taking the frequencies into account**.

$\bar{x} =$?

$\sigma^2 =$?

$\sigma =$?

You ABSOLUTELY must check this with the σx you can get on the calculator directly.

Test Your Understanding

May 2013 (R) Q3

An agriculturalist is studying the yields, y kg, from tomato plants. The data from a random sample of 70 tomato plants are summarised below.

Yield (y kg)	**Frequency** (f)	**Yield midpoint** (x kg)
$0 \leq y < 5$	16	2.5
$5 \leq y < 10$	24	7.5
$10 \leq y < 15$	14	12.5
$15 \leq y < 25$	12	20
$25 \leq y < 35$	4	30

(You may use $\sum fx = 755$ and $\sum fx^2 = 12\,037.5$)

(*c*) Estimate the mean and the standard deviation of the yields of the tomato plants. **(4)**

(c) ?

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.

Exercise 2E

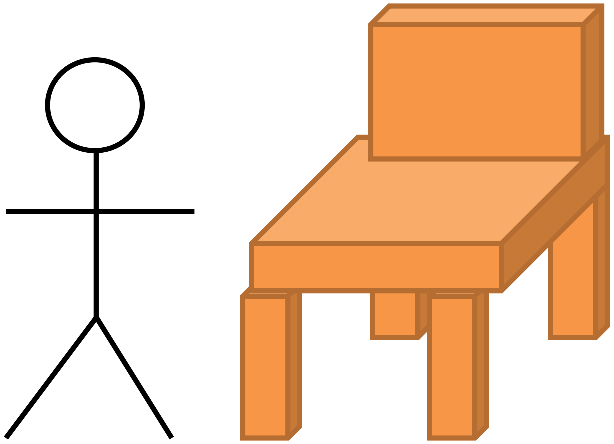
Pearson Statistics & Mechanics Year 1/AS
Pages 32-33

Coding

What do you reckon is the mean height of people in this room?

Now, stand on your chair, as per the instructions below.

INSTRUCTIONAL VIDEO



Is there an easy way to recalculate the mean based on your new heights? And the variance of your heights?

?

Starter

Suppose now after a bout of 'stretching you to your limits', you're now all 3 times your original height.

What do you think happens to the **standard deviation** of your heights?

?

What do you think happens to the **variance** of your heights?

?

Extension Question: Can you prove the latter using the formula for variance?

?

Rules of 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
$y = x + 10$	?	?
$y = 3x$	?	?
$y = 2x - 5$	?	?

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.

Rules of coding

Coding	Effect on $\bar{x}$	Effect on σ
$y = x + 10$		
$y = 3x$		
$y = 2x - 5$		

The point of coding

Cost x of diamond ring (£)

£1010	£1020	£1030	£1040	£1050

We 'code' our variable using the following:

$$y = \frac{x - 1000}{10}$$

New values y:

?

Standard deviation of y (σ_y): ?

therefore…

Standard deviation of x (σ_x): ?

The jist of coding: We want to find the mean/standard deviation of a variable. We transform the values, using some rule, to make them simpler. We can then more easily calculate the mean/standard deviation of the 'coded' data, and from this we can then determine what the mean/standard deviation would have been for the original uncoded data.

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

Example Exam Question

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)**
(*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)**
(*e*) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data. **(1)**

Suppose we've worked all these out already.

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(*f*) Without further calculations, explain the effect this would have on each of the estimates found in parts (*a*), (*b*), (*c*) and (*d*). **(3)**

f)

?

Exercise 2F

Pearson Statistics & Mechanics Year 1/AS

Pages 34-36

Chapter 2 Summary

I have a list of 30 heights in the class. What item do I use for:

- Q_1? ?
- Q_2? ?
- Q_3? ?

For the following grouped frequency table, calculate:

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

a) The estimate mean: ?

b) The estimate median: ?

c) The estimate variance: ?
(you're given $\Sigma f h^2 = 67.8125$)

Chapter 2 Summary

What is the standard deviation of the following lengths: 1cm, 2cm, 3cm

?

The mean of a variable x is 11 and the variance 4.

The variable is coded using $y = \frac{x+10}{3}$. What is:

a) The mean of y? ?

b) The variance of y? ?

A variable x is coded using $y = 4x - 5$.
For this new variable y, the mean is 15 and the standard deviation 8.
What is:

a) The mean of the original data? ?

b) The standard deviation of the original data? ?