Averages and Measures of Spread

Note Title 15/10/2008

We will consider 3 types of average (measures of location):

mode, median and mean

And 3 measures of spread (dispersion):

range, interquartile range, standard deviation

Mode and Range

The mode is the most common value.
If we have a grouped frequency distribution, we use the MODAL CLASS, which is the class with the greatest Frequency Density (ie the tallest bar on a histogram)
If a distribution has two modes, it is BIMODAL

The range is (the greatest value − the least value) (ie, it is a single number)

Median and Interquartile Range

We need to consider how to find these for :-

- individual data items
- a frequency distribution
- a grouped frequency distribution.

Individual data items (The Edexcel Way!)

If there are n items, work out $\frac{1}{2}n$ (for the median)
or $\frac{1}{4}n$ + $\frac{3}{4}n$ (for the quartiles)

If the result is not a whole number, round it UP (always) and use that value

If the result is a whole number (x, say) use the mean of the x^{th} and $(x+1)^{th}$ values.

Example 1: The reaction times (in seconds) of 33 students were recorded in a stem and leaf diagram as follows:

Reaction Times Key 1|3 means 0.13s

1	3 4 4	(3)
1	5 6 7 8 9	(5)
2	0 1 2 3 3 3 4 4 4 Q_2	(9)
2	5 6 6 6 7 7 8	
3	0 1 2 2 3 4	
3	5 8	
4		
4	9	

Median : $\frac{1}{2}$ of $33 = 16.5$

round up so median is 17^{th} value

So Median $(Q_2) = \underline{\underline{0.24s}}$

Lower Quartile (Q_1) : $\frac{1}{4}$ of $33 = 8.25$

round up so Q_1 is 9^{th} value

$Q_1 = \underline{\underline{0.20s}}$

Upper Quartile (Q_3) : $\frac{3}{4}$ of $33 = 24.75$

so Q_3 is 25^{th} value

$Q_3 = \underline{\underline{0.30s}}$

Interquartile Range $= Q_3 - Q_1 = \underline{\underline{0.10s}}$

$\uparrow Q_1$ $\uparrow Q_2$ $\uparrow Q_3$

Median : $\frac{1}{2}$ of $12 = 6$

WHOLE NUMBER so median is mean of 6^{th} & 7^{th} values

$$Q_2 = \underline{\underline{49.5g}}$$

Q_1 : Halfway between 3^{rd} & 4^{th} value : 45g

Q_3 : Halfway between 9^{th} & 10^{th} value : 62g

$$IQR = Q_3 - Q_1 = \underline{\underline{17g}}$$

Frequency Table

We use the same method as above, imagining the values in the table as a long list.

Example: The table shows the number of children living in each house on a certain housing development.

Number of children	Frequency	Cum Frequency
0	17	17
1	19	36
2	39	75
3	28	103
4	21	124
5	15	139
6	9	148
7	2	150

Find the median and quartiles of the number of children.

We imagine a long list of values:

$\underbrace{0\ 0\ 0 \ldots\ldots 0}_{17 \text{ '0's}}\ \underbrace{1\ 1\ 1 \ldots\ldots 1}_{19 \text{ '1's}}\ 2\ 2 \ldots\ldots$

Median: $\frac{1}{2}$ of $150 = 75$ so between 75^{th} + 76^{th} values

75^{th} value is the last '2'
76^{th} value is the first '3'

$Q_2 = 2.5$ children

Q_1: $\frac{1}{4}$ of $150 = 37.5$ so Q_1 is 38^{th} value

$Q_1 =$ 2 children

Q_3: $\frac{3}{4}$ of $150 = 112.5$ so Q_3 is 113^{th} value

$Q_3 =$ 4 children

p48 Ex 3A Q 2ab, 5, 9 HWK

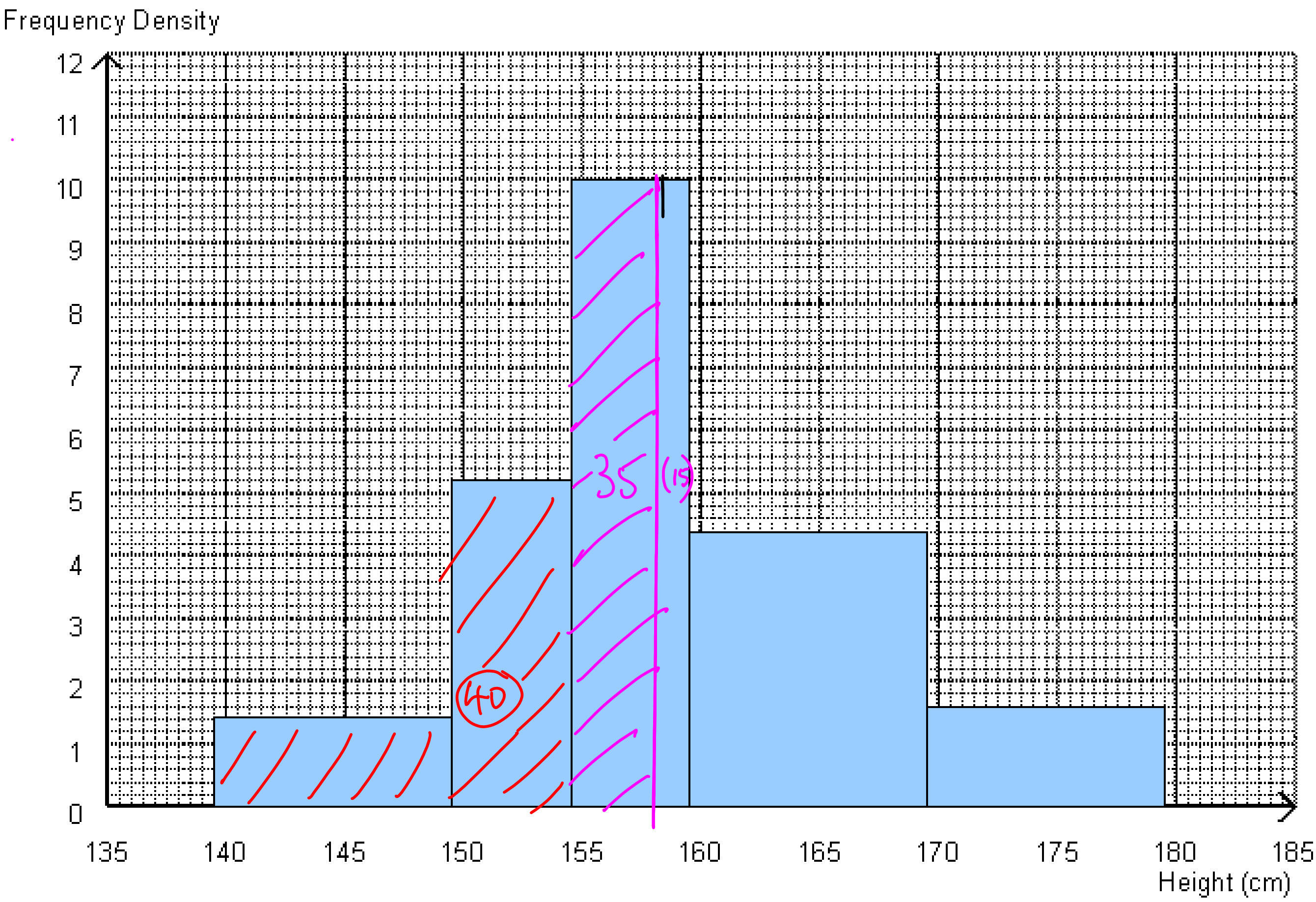
Grouped Frequency Table

In this case we do not know each individual data value. So we can only estimate the median. The method is different to the above :-

- We do not round up or use 'half-way values' – we just use $\frac{1}{2}n$, $\frac{1}{4}n$ or $\frac{3}{4}n$ as it stands
- We use LINEAR INTERPOLATION as in the example on the following page.

Example: The table below shows the heights (in cm) of 150 6th Form girls. Find the median and interquartile range.

Height	Frequency	Class Boundaries	Class Width	Cum Frequency
140 - 149	14	$139.5 \le x < 149.5$	10	14
150 - 154	26	$149.5 \le x < 154.5$	5	40
155 - 159	50	$154.5 \le x < 159.5$	5	90
160 - 169	44	$159.5 \le x < 169.5$	10	134
170 - 179	16	$169.5 \le x < 179.5$	10	150



We require $\frac{150}{2} = 75$ values below the median

So median is in the $154.5 - 159.5$ class

We require $75 - 40 = 35$ values out of the 50 in this class to be below the median.

So the median is $\frac{35}{50}$ of the way across this class

So median is at $154.5 + \left(\frac{35}{50} \times 5\right)$

$= 154.5 + 3.5$

$= \underline{\underline{158 \text{ cm}}}$

Lower Quartile:-

We require $\frac{150}{4} = 37.5$ values below Q_1

So Q_1 lies in the $149.5 - 154.5$ class.

We require $37.5 - 14 = 23.5$ of the 26 values to be below Q_1

So $Q_1 = 149.5 + \left(\frac{23.5}{26} \times 5\right)$

$= \underline{\underline{154.0 \text{ cm}}}$ (1 dp)

Upper Quartile :-

We require $\frac{3}{4} \times 150 = 112.5$ values below Q_3

So Q_3 lies in the $159.5 - 169.5$ class

So we need $112.5 - 90 = 22.5$ of the 44 values to be below Q_3

So $Q_3 = 159.5 + \left(\frac{22.5}{44} \times 10\right)$

$= \underline{\underline{164.6 \text{ cm}}}$ (1 dp)

$IQR = Q_3 - Q_1 = \underline{\underline{10.6 \text{ cm}}}$

Ex 3A Q 1(c), 2(c), 8(a)

Mean and Standard Deviation

Individual Data Items

The mean is written $\bar{x}$. If we have n values $x_1, x_2 \ldots x_n$, then

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \left(\text{often written } \frac{\sum x}{n}\right)$$

The standard deviation is a measure of spread – it looks at how far away each value is from the mean. The theory is as follows:–

1. For each x_i, calculate $x_i - \bar{x}$ to see how far x_i is from the mean.
2. Square each of these: $(x_i - \bar{x})^2$, so that they are all positive.
3. Find the mean of the values from step 2:

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

[This value is called the VARIANCE of the data]

4. Find the square root of the variance (to compensate for the squaring in step 2.

So the standard deviation $= \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n}}$.

However, this can be manipulated algebraically into a form which is usually easier to use.

$$\boxed{\text{This is standard deviation} = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}}$$

(see p57 of S1 for the algebra involved)

<u>Example</u> Find the mean and standard deviation of the following values.

1 4 8 9 10 12 12 15 19 20

$\sum x = 110$ so $\bar{x} = \frac{110}{10} = \underline{\underline{11}}$

$\sum x^2 = 1536$ so $s.d = \sqrt{\frac{1536}{10} - 11^2}$

$= \underline{\underline{5.71}}$ (3 sf)

Ex 4A Q 3(a), 4(a), 5, 11

Grouped (or Ungrouped) Frequency Distributions

The mean of a frequency distribution is

$$\boxed{\bar{x} = \frac{\sum fx}{\sum f}}$$

(where x is the <u>midpoint</u> of each class if the distribution is grouped)

So the formula for standard deviation becomes

$$\boxed{s.d = \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}}$$

not $(fx)^2$!

Height	Frequency (f)	Class Midpoints (x)	fx	fx^2
140 - 149	14	144.5	2023	292323.5
150 - 154	26	152	3952	600704
155 - 159	50	157	7850	1232450
160 - 169	44	164.5	7238	1190651
170 - 179	16	174.5	2792	487204

$\sum f = 150$ $\quad\sum fx = 23855$ $\quad\sum fx^2 =$

$$\text{Mean} \quad \bar{x} = \frac{\sum fx}{\sum f} = \frac{23855}{150} = 159.0\bar{3} \text{ cm}$$

$$SD = \sqrt{\frac{3803332.5}{150} - 159.0\bar{3}^2}$$

$$= 7.9968\ldots$$

$$= \underline{\underline{8.00}} \text{ cm } (3 \text{ sf})$$

p65 Ex 4A Q 3(b)(c), 7, 9

Using 'Coding' to make the numbers smaller

This is based on the principles that:

If we add a constant c to each x_i,

- the mean $\bar{x}$ is increased by c
- the standard deviation σ_x does not change

If we multiply each x_i by a constant a

- the mean $\bar{x}$ is multiplied by a
- the standard deviation σ_x is multiplied by a

So if we CODE the x_i values using a formula

$$\boxed{y_i = \frac{x_i - c}{a}}$$

(we choose c and a to make the 'y_i's nice simple numbers)

then the uncoding formula is

$$x_i = a y_i + c$$

So

$$\boxed{\begin{aligned} \bar{x} &= a\bar{y} + c \\ \sigma_x &= a\sigma_y \end{aligned}}$$

Example: The table below shows the heights (in cm) of 150 6th Form girls. Use coding to find the mean and standard deviation.

$x = y \times 2.5 + 157$

Height	Frequency (f)	Class Midpoint (x)	$y = \frac{x - 157}{2.5}$	fy	fy^2 ← NOT $(fy)^2$
140 – 149	14	144.5	−5	−70	350
150 – 154	26	152	−2	−52	104
155 – 159	50	157	0	0	0
160 – 169	44	164.5	3	132	396
170 – 179	16	174.5	7	112	784
	$\Sigma f = 150$			$\Sigma fy = 122$	$\Sigma fy^2 = 1634$

$$\bar{y} = \frac{122}{150} = 0.81\dot{3}$$

$$\sigma_y = \sqrt{\frac{1634}{150} - 0.81\dot{3}^2} = 3.1988$$

So $\bar{x} = 0.81\dot{3} \times 2.5 + 157$

$= \underline{\underline{159.0\dot{3}}}$

$\sigma_x = 3.1988 \times 2.5$

$= 7.9968\ldots$

$= \underline{\underline{8.00}}$ (3sf)

p66 Ex 4A Q 8, 10 for HWK