Statistics Activity 4

‘How do you know if you are really unwell?’

When you are feeling unwell and visit the doctor, one of the first things the doctor does is to take your temperature. Your temperature is then compared with the accepted ‘normal body temperature’ of 98.6°F (37.0°C) and conclusions drawn. This figure was first agreed upon over 100 years ago and recent evidence suggests that this figure may not be representative of the world we now live in. In this activity we look at the some of the evidence that apparently supports this view.

Whenever possible, data analysis should be based on real data, particularly if we are trying to draw conclusions about the real world. Consequently this activity is based on an article published in the *Journal of the American Medical Association* that examined whether the true mean body temperature is 98.6°F (37.0°C). The data provided appears to suggest that ‘normal body temperature’ is slightly lower than previously thought, at 98.2°F (36.8°C). In addition, students are encouraged to ask further questions such as ‘At what temperature should we consider someone’s temperature to be abnormal?’ and ‘Is there a significant difference between the normal temperature of males and females?’

This activity gives the students an introduction to how to summarise continuous data. It also looks at the effect of changing the size of the class interval on the summary statistics. This dataset will be used again in a subsequent activity.

Features of this activity

- Exploration of a real, everyday problem
- The use of a real dataset
- The six key questions you should ask when using someone else’s data
- Summarising continuous data
- The effect of the size of the class interval on these summary statistics
- A discussion of the difficulties of drawing conclusions for the population based on a sample

Materials and preparation

- Introductory video (to be supplied)
- Introductory PowerPoint (bodytemperature.ppt)
- Notebook file (bodytemperature) (This is a Smartboard file.)
- Data sheet

How do you know if you are really unwell?

Introducing the problem

Introduce the problem by asking ‘How do you know if you are really unwell?’ This can be through group discussion or by using either the PowerPoint file or Notebook file (Smartboard).

You need to make sure that this includes ‘taking your temperature’.

You can then follow up with questions such as ‘Why does the doctor take your temperature?’ ‘How does this help the doctor to decide if you are unwell?’ ‘What is normal body temperature?’

Introducing the dataset

Introduce the activity by emphasising the need for reliable data if we are going to draw any meaningful conclusions.

Stress that when using someone else’s data it is important to ask six questions to establish the credibility and relevance of the data. (This is covered in the introductory notes for this pathway.)

This activity is based on data published in the *Journal of the American Medical Association* that examined whether the true mean body temperature is 98.6°F (37.0°C). This data appears to suggest that ‘normal body temperature’ is slightly lower than previously thought, at 98.2°F (36.8°C).

Now display the original data, shown below in a stem and leaf plot. Stress that the data has been rounded to 1dp. Ask your students what this actually shows.

The dataset

Male	35	79	(2)
	36	112222333334444	(15)
	36	566666677777777888888999999	(27)
	37	0000001111122223344	(20)
	37	5	(1)
			(65)
Female	35	89	(2)
	36	02234	(5)
	36	556666667777778888888888999999	(29)
	37	000011111111111111222333344	(26)
	37	78	(2)
	38	2	(1)
			(65)

Key 35 | 8 represents 35.8

Discuss the data.

Prompt with questions such as: ‘What do you notice about this data?’

‘Is there any difference between genders?’

‘Clearly the data has been rounded to 1dp. Why is that?’

Try to bring out that this is continuous data that has been rounded. Establish that a temperature (t) of 37.1°C actually means 37.05 ≤ t < 37.15. The difference between the upper and lower bounds is known as the 'class interval' or 'bin width'. In this case it is 0.1.

Make sure everyone understands this before moving on.

Ask your students how else this data might be presented, especially as there is a lot of data and much of it is repeated. For example, 13 females had a temperature of 37.1°C.

Try to get to a frequency diagram as shown below.

Male data in a frequency table

(Class interval width 0.1)

Temperature (°C) True class interval	Mid value Temperature(°C) (x)	Frequency (f)
35.65 ≦ t < 35.75	35.7	1
35.85 ≦ t < 35.95	35.9	1
36.05 ≦ t <36.15	36.1	2
36.15 ≦ t < 36.25	36.2	4
36.25 ≦ t < 36.35	36.3	5
36.35 ≦ t < 36.45	36.4	5
36.45 ≦ t 36.55	36.5	1
36.55 ≦ t < 36.65	36.6	6
36.65 ≦ t < 36.75	36.7	8
36.75 ≦ t < 36.85	36.8	6
36.85 ≦ t < 36.95	36.9	6
36.95 ≦ t < 37.05	37.0	6
37.05 ≦ t < 37.15	37.1	5
37.15 ≦ t < 37.25	37.2	4
37.25 ≦ t < 37.35	37.3	2
37.35 ≦ t < 37.45	37.4	2
37.45 ≦ t < 37.55	37.5	1
	Total	65

Estimating the mean - Female results

(Class interval width 0.1)

Temperature (°C) True class interval	Mid value Temperature(°C) (x)	Frequency (f)
35.75 ≦ t < 35.85	35.8	1
35.85 ≦ t < 35.95	35.9	1
35.95 ≦ t < 36.05	36.0	1
36.15 ≦ t < 36.25	36.2	2
36.25 ≦ t < 36.35	36.3	1
36.35 ≦ t < 36.45	36.4	1
36.45 ≦ t < 36.55	36.5	2
36.55 ≦ t < 36.65	36.6	6
36.65 ≦ t < 36.75	36.7	6
36.75 ≦ t < 36.85	36.8	9
36.85 ≦ t < 36.95	36.9	6
36.95 ≦ t < 37.05	37.0	4
37.05 ≦ t < 37.15	37.1	13
37.15 ≦ t < 37.25	37.2	3
37.25 ≦ t < 37.35	37.3	4
37.35 ≦ t < 37.45	37.4	2
37.65 ≦ t < 37.75	37.7	1
37.75 ≦ t < 37.85	37.8	1
38.15 ≦ t < 38.25	38.2	1
	Total	65

Summarising continuous data from a frequency diagram

1. Estimating the mean

'Discuss with your partner, how you might do this and why it is an "estimate" of the mean.'

Prompt as necessary until your students all understand what they need to do. The correct solutions in this case are shown in the following boxes.

Estimating the mean - Male results

(Class interval width 0.1)

Temperature (°C) True class interval	Mid value Temperature(°C) (x)	Frequency (f)	fx
35.65 ≦ t < 35.75	35.7	1	35.7
35.85 ≦ t < 35.95	35.9	1	35.9
36.05 ≦ t <36.15	36.1	2	72.2
36.15 ≦ t < 36.25	36.2	4	144.8
36.25 ≦ t < 36.35	36.3	5	181.5
36.35 ≦ t < 36.45	36.4	5	182.0
36.45 ≦ t 36.55	36.5	1	36.5
36.55 ≦ t < 36.65	36.6	6	219.6
36.65 ≦ t < 36.75	36.7	8	293.6
36.75 ≦ t < 36.85	36.8	6	220.8
36.85 ≦ t < 36.95	36.9	6	221.4
36.95 ≦ t < 37.05	37.0	6	222.0
37.05 ≦ t < 37.15	37.1	5	185.5
37.15 ≦ t < 37.25	37.2	4	148.8
37.25 ≦ t < 37.35	37.3	2	74.6
37.35 ≦ t < 37.45	37.4	2	74.8
37.45 ≦ t < 37.55	37.5	1	37.5
	Total	65	2387.2

Estimated mean $= \frac{2387.2}{65} = 36.7$

Estimating the mean - Female results

(Class interval width 0.1)

Temperature (°C) True class interval	Mid value Temperature(°C) (x)	Frequency (f)	fx
35.75 ≦ t < 35.85	35.8	1	35.8
35.85 ≦ t < 35.95	35.9	1	35.9
35.95 ≦ t < 36.05	36.0	1	36.0
36.15 ≦ t < 36.25	36.2	2	72.4
36.25 ≦ t < 36.35	36.3	1	36.3
36.35 ≦ t < 36.45	36.4	1	36.4
36.45 ≦ t < 36.55	36.5	2	73.0
36.55 ≦ t < 36.65	36.6	6	219.6
36.65 ≦ t < 36.75	36.7	6	220.2
36.75 ≦ t < 36.85	36.8	9	331.2
36.85 ≦ t < 36.95	36.9	6	221.4
36.95 ≦ t < 37.05	37.0	4	148.0
37.05 ≦ t < 37.15	37.1	13	482.3
37.15 ≦ t < 37.25	37.2	3	111.6
37.25 ≦ t < 37.35	37.3	4	149.2
37.35 ≦ t < 37.45	37.4	2	74.8
37.65 ≦ t < 37.75	37.7	1	37.7
37.75 ≦ t < 37.85	37.8	1	37.8
38.15 ≦ t < 38.25	38.2	1	38.2
	Total	65	2397.8

Estimated mean $= \frac{2397.8}{65} = 36.9$

2. Does the size of the class interval matter?

'What happens if we make the class interval bigger?'

Clearly there will be fewer calculations and this will make life easier, **but** will the results still be reasonable?

'For example, what happens if we make the class interval 0.2?'

Let your students try it for themselves, but it is probably a good idea to make sure to start off together on the board.

The results for a class interval of 0.2 are shown in the boxes following.

As you can see, there is little difference in this case.

Why is this?

What happens if we make the bin wider still?

This will be covered later but it depends on the shape of your original data.

Estimating the mean - Male results

(Class interval width 0.2)

Temperature (°C) True class interval	Mid value Temperature(°C) (x)	Frequency (f)	fx
35.65 ≦ t < 35.85	35.75	1	35.75
35.85 ≦ t < 36.05	35.95	1	35.95
36.05 ≦ t <36.25	36.15	6	216.9
36.25 ≦ t < 36.45	36.35	10	363.5
36.45 ≦ t < 36.65	36.55	7	255.85
36.65 ≦ t < 38.85	36.75	14	514.5
36.85 ≦ t < 37.05	36.95	12	443.4
37.05 ≦ t < 37.25	37.15	9	334.35
37.25 ≦ t < 37.45	37.35	4	149.4
37.45 ≦ t < 37.65	37.55	1	37.55
	Total	65	2387.15

Estimated mean = $\frac{2387.15}{65}$ = 36.7

Estimating the mean - Female results

(Class interval width 0.2)

Temperature (°C) True class interval	Mid value Temperature(°C) (x)	Frequency (f)	fx
35.75 ≦ t < 35.95	35.85	2	71.7
35.95 ≦ t < 36.15	36.05	1	36.05
36.15 ≦ t < 36.35	36.25	3	108.75
36.35 ≦ t < 36.55	36.45	3	109.35
36.55 ≦ t < 36.75	36.65	12	439.8
36.75 ≦ t < 36.95	36.85	15	552.75
36.95 ≦ t < 37.15	37.05	17	629.85
37.15 ≦ t < 37.35	37.25	7	260.75
37.35 ≦ t < 37.55	37.45	2	74.9
37.55 ≦ t < 37.75	37.65	1	37.65
37.75 ≦ t < 37.95	37.85	1	37.85
37.95 ≦ t < 38.15	38.05	0	0
38.15 ≦ t < 38.35	38.25	1	38.25
	Total	65	2397.65

Estimated mean = $\frac{2397.65}{65}$ = 36.9

3. Estimating the median

'Discuss with your partner how we might be able to estimate the median when the data is given in a frequency table.'

Most of your students will have covered this at GCSE but may need prompting to get to a cumulative frequency table and hence a cumulative frequency graph.

To make life easier, it is a good idea to continue with a class interval of 0.2.

Make sure your students do understand where the third column comes from and what it actually means. For example, ask: 'How many female students have a temperature below 36.5°C?'

Make sure your students understand that the last entry in the cumulative frequency column should be the same as the total of the frequencies.

The results for both male and female students are shown in the following boxes.

Estimating the median - Male results

(Class interval width 0.2)

Temperature (°C)True class interval	Frequency (f)	Cumulative frequency
35.65 ≦ t < 35.85	1	1
35.85 ≦ t < 36.05	1	2
36.05 ≦ t <36.25	6	8
36.25 ≦ t < 36.45	10	18
36.45 ≦ t < 36.65	7	25
36.65 ≦ t < 38.85	14	39
36.85 ≦ t < 37.05	12	51
37.05 ≦ t < 37.25	9	60
37.25 ≦ t < 37.45	4	64
37.45 ≦ t < 37.65	1	65
	65	

To estimate the median now draw a cumulative frequency graph.

Estimating the median - Female results

(Class interval width 0.2)

Temperature (°C)True class interval	Frequency (f)	Cumulative frequency
35.75 ≦ t < 35.95	2	2
35.95 ≦ t < 36.15	1	3
36.15 ≦ t < 36.35	3	6
36.35 ≦ t < 36.55	3	9
36.55 ≦ t < 36.75	12	21
36.75 ≦ t < 36.95	15	36
36.95 ≦ t < 37.15	17	53
37.15 ≦ t < 37.35	7	60
37.35 ≦ t < 37.55	2	62
37.55 ≦ t < 37.75	1	63
37.75 ≦ t < 37.95	1	64
37.95 ≦ t < 38.15	0	64
38.15 ≦ t < 38.35	1	65
	65	

4. Plotting a cumulative frequency graph and estimating the median

As most of your students will have covered this at GCSE, they should be prompted to explain how to proceed. Again it is worth starting it on the board to make sure of the following points.

a) The horizontal axis need not start at zero. (Explain how this is shown.)

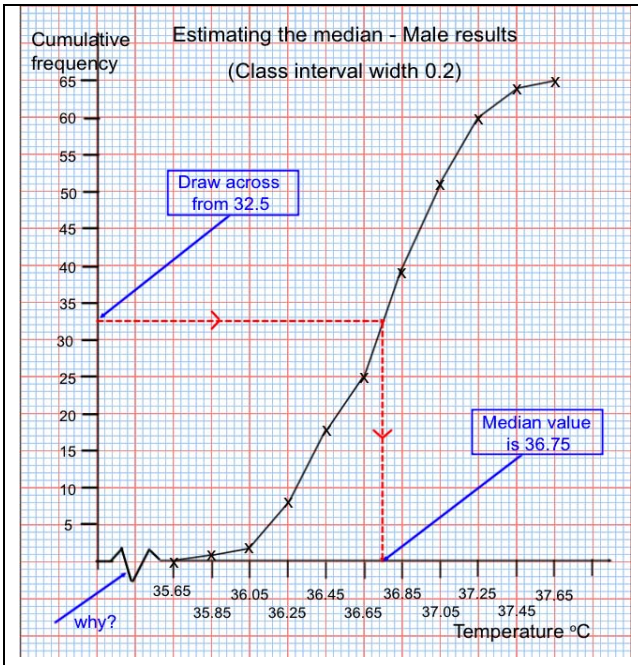
b) The data points are plotted at the end of each class interval.

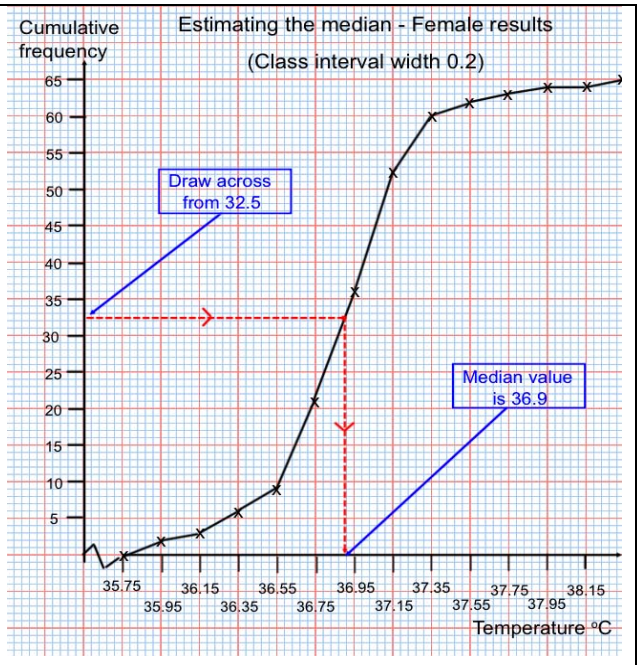
c) With real data, the choice of scale may be difficult.

d) The graph should be large enough to enable the students to make reasonable estimates.

e) To find an estimate for the median, first you must decide where to draw the horizontal line. As this is an estimate, simply dividing the frequency by two is usually accurate enough. So in this case, the horizontal line starts at 32.5 as shown in the graphs following.

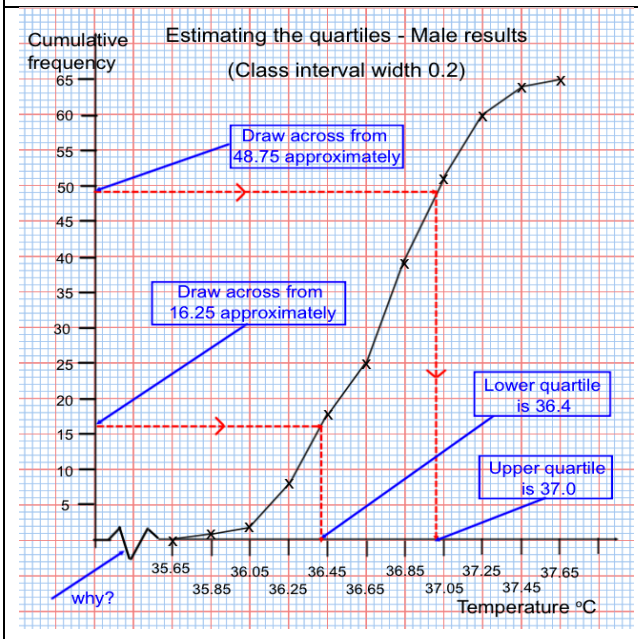
f) Encourage your students to mark in their construction lines clearly on their graph.

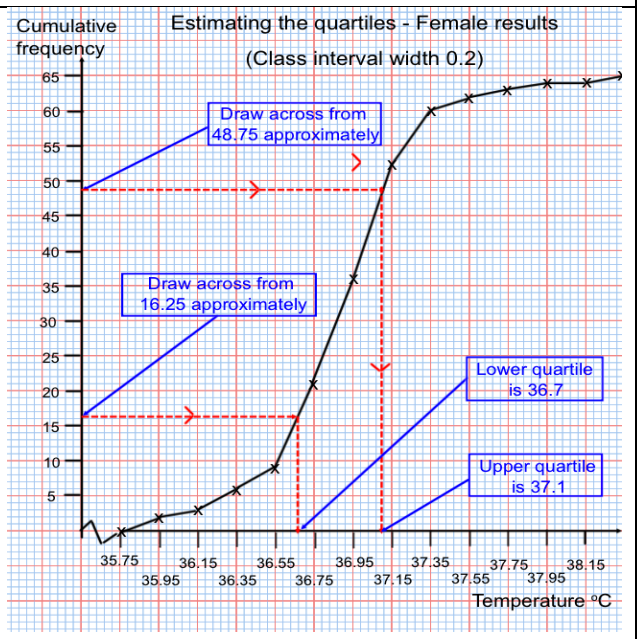




5. Estimating the quartiles

These estimates follow in a similar way by drawing further horizontal lines at 16.25 (0.25 x 65) and 48.75 (0.75 x 65). Clearly this is difficult to do accurately but it doesn't really matter as it is only meant to be an estimate.



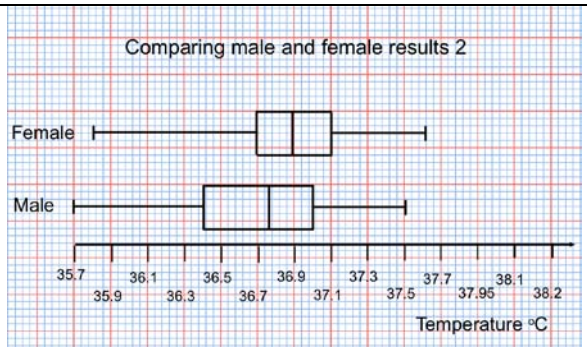


6. Compiling our statistics

Finally we begin to think about answering our question, 'How do you know if you are really unwell?' To do this, construct a summary chart of everything we have found out so far (see following). It is also worthwhile to plot two box plots using the same scale, one underneath the other, to enable the students to make a comparison between the male and female results. You may need to remind your students how to find any outliers. In this case, we can suggest that if you are an outlier, you may be considered 'unwell'.

Comparing male and female results 1

Statistic	Male	Female
Mean	36.7	36.9
Mode	36.7	37.1
Minimum	35.7	35.8
Lower quartile	36.4	36.7
Median	36.75	36.9
Upper quartile	37.0	37.1
Maximum	37.5	38.2
Range	1.8	1.3
IQR	0.6	0.4
Low outliers	< 35.5	< 36.1
High outliers	> 37.9	> 37.3



7. Analysing the results

Now ask your students to think about what the summary statistics indicate. Encourage them to share their thoughts with their partner before opening up the discussion to the group.

In particular, ask them to think about how they would answer the following:

‘What is the average temperature of a male?’

‘What is the average temperature of a female?’

‘Are there any significant differences between the male and female temperatures?’

‘How many students might be unwell, based on this data?’

‘Could this data be used to make any claims about the student population in general?’