MEDIAN, QUATILES, PERCENTILES!

HOMEWORK:

Data given as list + more: page 55-56-2F: all (very important)
Data given discrete-non grouped: page 57-58-2G: all (very important)
Data given as grouped (discrete/continuous): page 65-66-2I: all (very important)

This is due next class, Tuesday, May 6th

THE BIG PITURE!

- Measurements of Central Tendencies:
  - ☑ mode or modal group
  - ☑ mean
  - ☐ median
- Measurements of Spread:
  - ☐ Range
  - ☐ IQR
  - ☐ Std

Definitions

- Percentiles: (no longer part of your curriculum):
- **Median:** the middle value when data is arranged in size order. It does not take into account ALL the elements (or your x's) of your sample

$= 2^{nd}$ Quartile $2 \times \frac{1}{4} = \frac{1}{2}$

- **1st Quartile**: It is the median of the first half of the data.

1 quater of your data $= \frac{1}{4}$

also called: LOWER QUARTILE

- **3rd Quartile**: It is the median of the 2nd half of the data.

3 quaters of your data $= \frac{3}{4}$

also called: UPPER QUARTILE

MEDIAN- Discrete-list

STEP 1:
Put data in increasing order

0 1 2 2 1 3 4 3 1 2 0 0 1 0 2 1 0 1 0 1

example 1:

0 0 0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2

3 3 4

$\frac{1+1}{2} = 1$

Example 2: 2 4 8 10 11

STEP 2: Find position of your median by using the quick formula (n+1)/2 (see next slide)

STEP 3: If the position has a decimal, then you "grab" those two terms of your data and calculate the average of the two
If the position is a whole number, just "grab" that term!

the position of median

$\frac{n+1}{2}$

example 1: $\frac{20+1}{2} = 10.5$ → 10th term, 11th terms and do the avg

10th term = 1
11th term = 1
∴ Median = $\frac{1+1}{2}$ = 1

example 2: $\frac{5+1}{2} = 3^{Rd}$ term

the 3rd term is 8, hence median = 8

Median & Quartiles-Discrete List

1st → 3rd

0 1 2 2 1 3 4 3 1 2 0 0 1 0 2 1 0 1 0 1

0 0 0 0 0 0 1 1 1 1 1 2 2 2 2 3 3 4

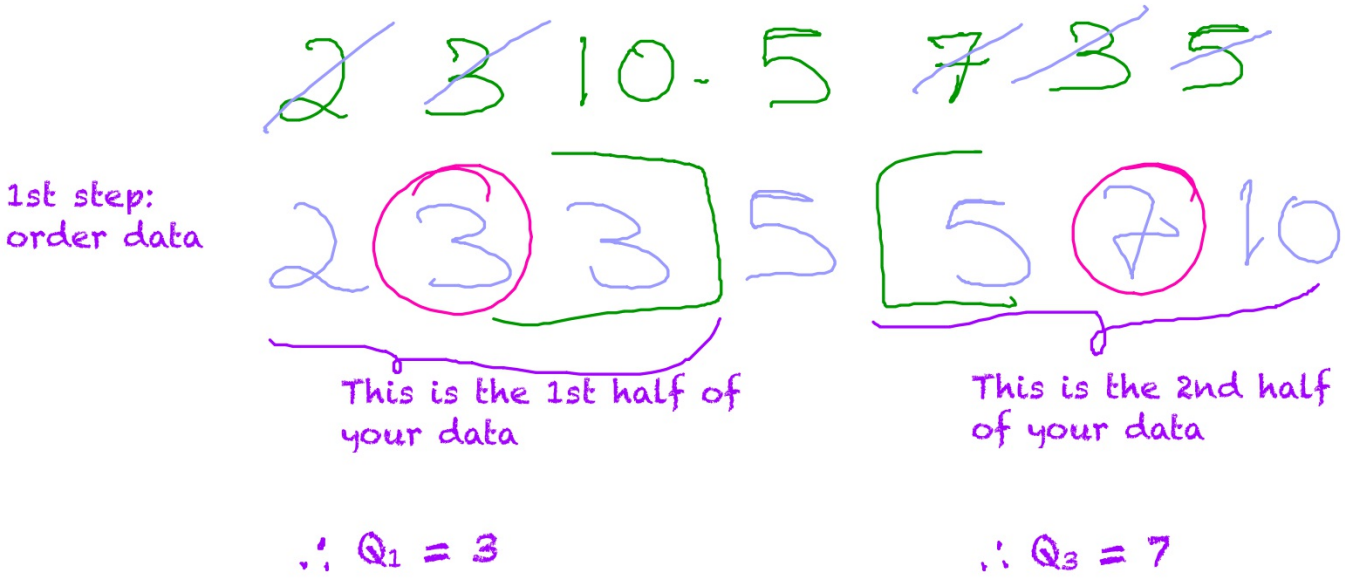
5th 6th 5th 6th

Remember: Q_2 is the median of 1st half of the data
Q_3 is the median of the 2nd hald of the data. Hence now n = 10

Finding positions ==> $\frac{10+1}{2} = 5.5$ hence I need to "grap" 5th term, 6th term → average

Therefore, for $Q_1 = \frac{0+0}{2} = 0$ and for $Q_3 = \frac{2+2}{2} = 2$

In the previous example, you could easily divide the data into 2 equal halfs, because n was an even number (i.e. 20), but what do you do if n = odd number?
Answer: Ignore the middle value (the median)



There is a quicker way to get the position of your quartiles, it is a shortcut that works 99.5% of the time.

To find position of Q_1, apply this formula: $\frac{n+1}{4}$

To find position of Q_3, apply this formula: $\frac{3(n+1)}{4}$

For example: let's say n=26

Then to get the position of Q_1, $\frac{26+1}{4} = 6.75$ so "grab" 6th term, 7th term

To get the position of Q_3, $\frac{3(26+1)}{4} = 20.25$ so grab 20th term, 21st term

And do the averages.

Going backward!

- If the median is 5, can you find the missing value a:

Example 1: 1st step: find position $\frac{n+1}{2} = \frac{7+1}{2} = 4^{th}$ term: a

2, 4, 5, a, 7, 8, 10

$\therefore a = 5$

Example 2: Solve for a if the median of the below data is 6.5

2, 4, 5, a, 7, 8, 8, 10

1st step: find position: $\frac{8+1}{2} = 4.5$ So now we know that the median is the average of 4th term: a and 5th term: 7

2ns step: solve for a: $\frac{a+7}{2} = 6.5$

$a = 6$

DIFFERENT FREQUENCIES

- Different frequencies:

1 Raw Frequency It is the frequency (the how many) of one category for your variable

2 Relative Frequency: It is the raw frequency, but relative to the total (Σf)
Similar to a %, but you do not multiply be a 100

Note: you will not see this in math studies tests.

3 CUMULATIVE Frequency It is the total (sum) up to one category, hence the name "cumulative"

Median and quartiles: discrete-freq. table

x	Raw f	Relative Frequency	Cumulative f
1	2	0.1	2
2	10	0.50	12
3	2	0.1	14
4	3	0.15	17
5	1	0.05	18
6	2	0.1	20
	$\Sigma f = 20$	Note: these should add up to 1	

Median or Q2: $\frac{n+1}{2} = 10.5$ → 10th term, 11th term

$$\text{Median} = \frac{2+2}{2} = 2$$

Q_1

$$\frac{n+1}{4} = 5.25 \begin{cases} \rightarrow 5^{th} \\ \rightarrow 6^{th} \end{cases}$$

$$Q_1 = \frac{2+2}{2} = 2$$

Q_3

$$\frac{3(n+1)}{4} = 15.75 \begin{cases} \rightarrow 15^{th} \\ \rightarrow 16^{th} \end{cases}$$

$$Q_3 = \frac{4+4}{2} = 4$$

MEDIAN-Discrete-Frequency Table

x	f	cumulative f
120	2	2
121	5	7
122	12	19
123	10	29
124	5	34
	$\sum f = 34$	

120 120 121 121 ... 121
122 122 ... 122

Median

$\frac{n+1}{2} = \frac{35}{2} = 17.5$

17th term, 18th term

Median $= \frac{122+122}{2} = 122$

1st quartile: $\frac{n+1}{4} \Rightarrow \frac{34+1}{4} = 8.75$ — 8th term 122, 9th term 122 — average

$Q_1 = \frac{122+122}{2} = \boxed{122}$

3rd quartile: $\frac{(n+1)3}{4} \Rightarrow \frac{(34+1)3}{4} = 26.25$ — 26th term 123, 27th term 123 — average

$Q_3 = \frac{123+123}{2} = \boxed{123}$

MEDIAN-Discrete-Grouped

Number of cars	Frequency	cumulative f.	Upper boundary
0 to 9	1	1	9.5
10 to 19	5	6	19.5
20 to 29	10	16	29.5
30 to 39	9	25	39.5
40 to 49	4	29	49.5
50 to 59	1	30	59

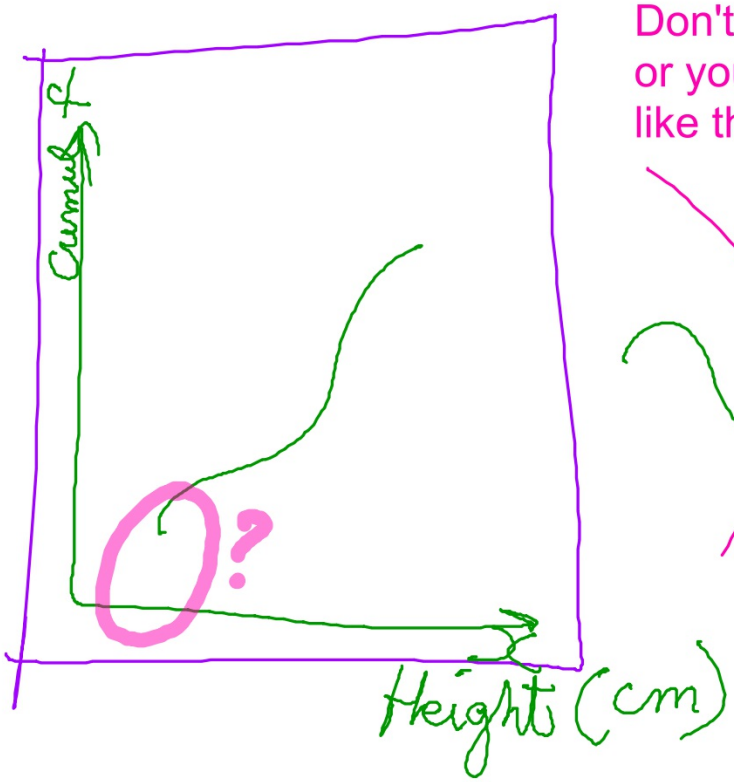
You can only find the median by drawing a:
"Cumulative Frequency Graph"

NOTE: Continuous data is easier than discrete-grouped "you do not have to worry about "correct" upper boundary.

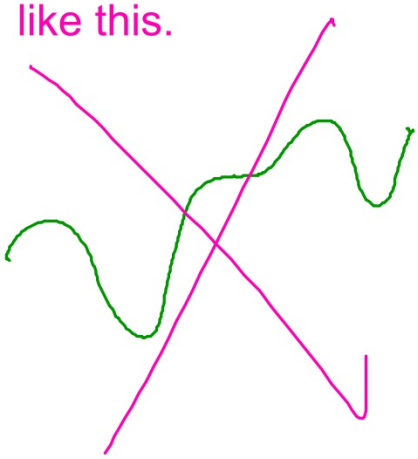
Median & Quartiles (grouped data)

Cumulative Frequency graph/curve

1. Get graph paper and draw your axis. Label them correctly. (Remember that you are only working with the 1st quadrant)
2. Plot your points.
   - ✓ Get correct points from your table-use:
     (Upper boundary, cum. freq)
3. Draw a smooth curve
4. Find the value of your variable corresponding to the middle term. (50%), the 1st quarter term (25%) and the 3rd quarter term (75%)



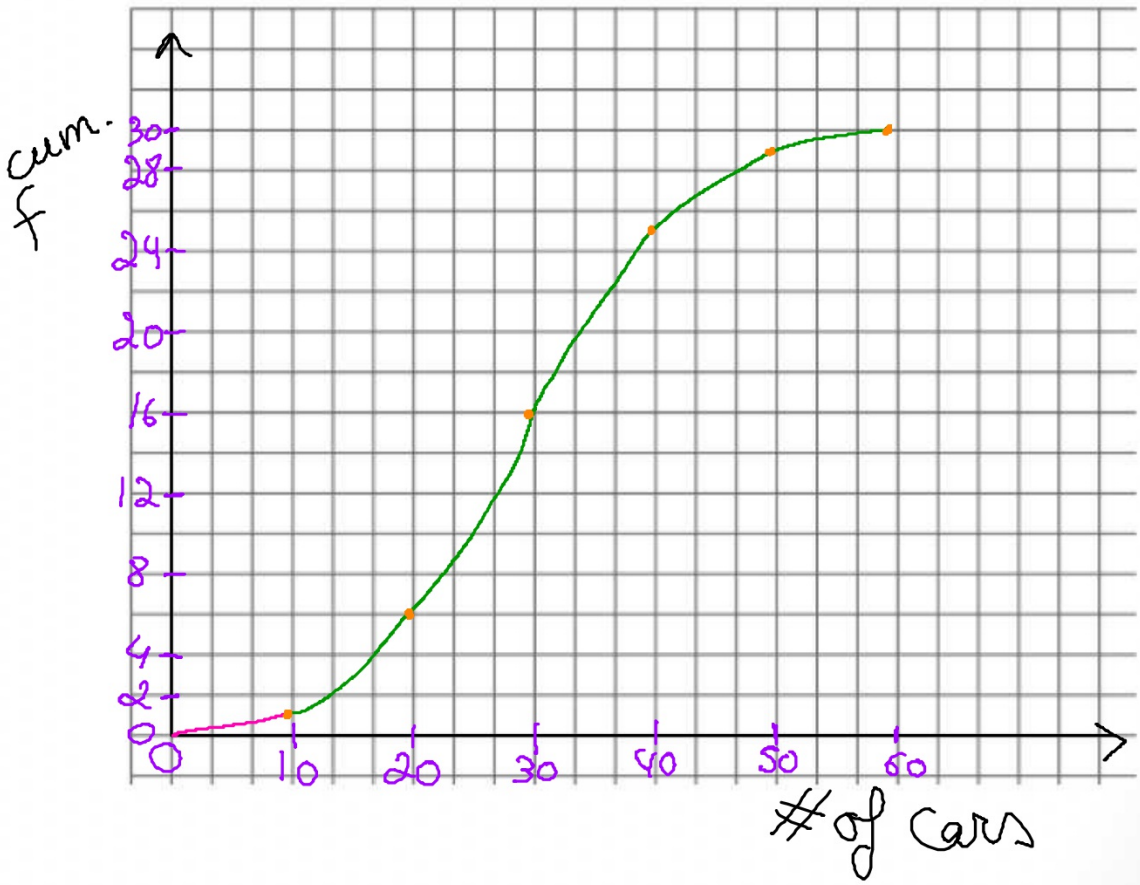
Don't use RAW f or you get something like this.



MEDIAN-Discrete-Grouped

Number of cars	Frequency	cumulative f.	Upper boundary	points
0 to 9	1	1	9.5	(9.5, 1)
10 to 19	5	6	19.5	(19.5, 6)
20 to 29	10	16	29.5	(29.5, 16)
30 to 39	9	25	39.5	(39.5, 25)
40 to 49	4	29	49.5	(49.5, 29)
50 to 59	1	30	59	(59, 30)

You cannot have a gap in your curve, which means you need to pick an upper boundary, the best boundary is the compromise between endpoints. This should remind you to whey you were drawing Histograms.



MEDIAN-Continuous- Grouped

- What do you think? How should we do this?

Since it is grouped, we must draw a:
cumulative frequency curve---> DUH!!

Remember: cum. freq. curves always go up and up!

NOTE: This time it is easier than with discrete-grouped. Why? Because we do not have to worry about finding the "correct" upper boundary.

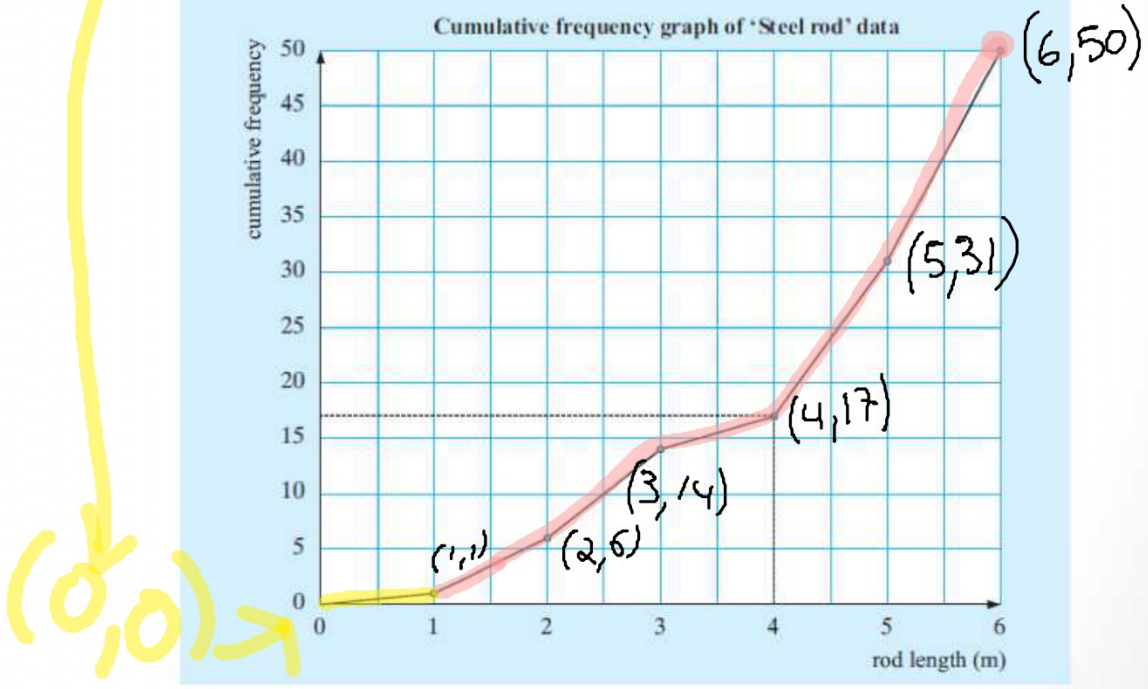
Median & Quartiles (grouped data)

Cumulative Frequency graph/curve

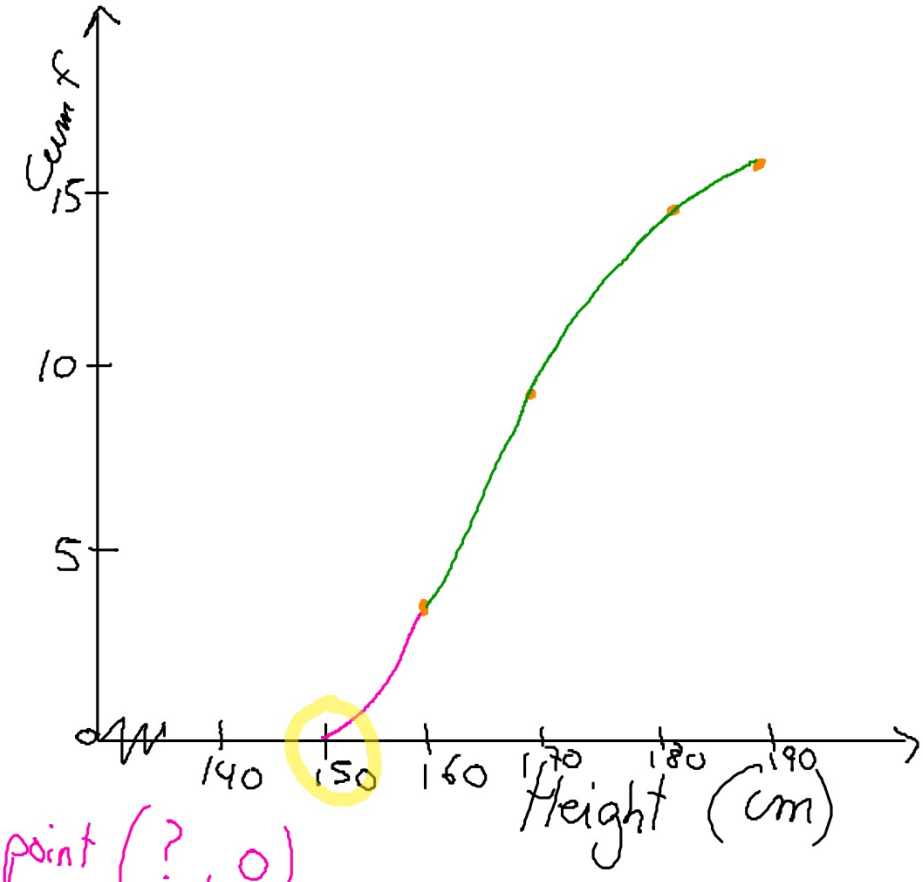
1st step: Find the points.....

Rod Length (metres)	*Tally*	*Frequency*	*Cumulative Frequency*	points (ℓ, c.f)
0 - 0.99 <1	|	1	1	(1, 1)
1.00 - <1.99 2	𝍸	5	6	(2, 6)
2.00 - <2.99 3	𝍸 |||	8	14	(3, 14)
3.00 - <3.99 4	|||	3	17	(4, 17)
4.00 - <4.99 5	𝍸 𝍸 ||||	14	31	(5, 31)
5.00 - <5.99 6	𝍸 𝍸 𝍸 ||||	19	50	(6, 50)
Total		50		

Rod length (metres)	*Upper End Point*	*Cumulative Frequency*
0 - 0.99	0.995	1
1 - 1.99	1.995	6
2 - 2.99	2.995	14
3 - 3.99	3.995	17
4 - 4.99	4.995	31
5 - 5.99	5.995	50



(cm) Heights	f	cf
$150 \le h < 160$	3	3
$160 \le h < 170$	6	9
$170 \le h < 180$	5	14
$180 \le h < 190$	2	16



last point (?, 0)

(150, 0)

Now, you are ready to:

Find the Median: 50 percentile

Find Quartiles:

25th percentile and 75th percentile

Q_1 Q_3

$n = 200$ $n = 120$

50% percentile $\Rightarrow$ at 100 | at 60

25% " $\Rightarrow$ at 50 | at 30

75% " $\Rightarrow$ at 150 | at 90

Rod length (metres)	Upper End Point	Cumulative Frequency	Points
0 - 0.99 < 1	0.995	1	(1 ,1)
1 - 1.99 < 2	1.995	6	(2, 6)
2 - 2.99 < 3	2.995	14	(3, 14)
3 - 3.99 < 4	3.995	17	(4, 17)
4 - 4.99 < 5	4.995	31	(5, 31)
5 - 5.99 < 6	5.995	50 n = 50	(6, 50)

To find the position or frequency of the Median

n times 50%
50 X 0.50 = 25

Median = 4.55 ±0.05

To find the position or frequency of Q_1

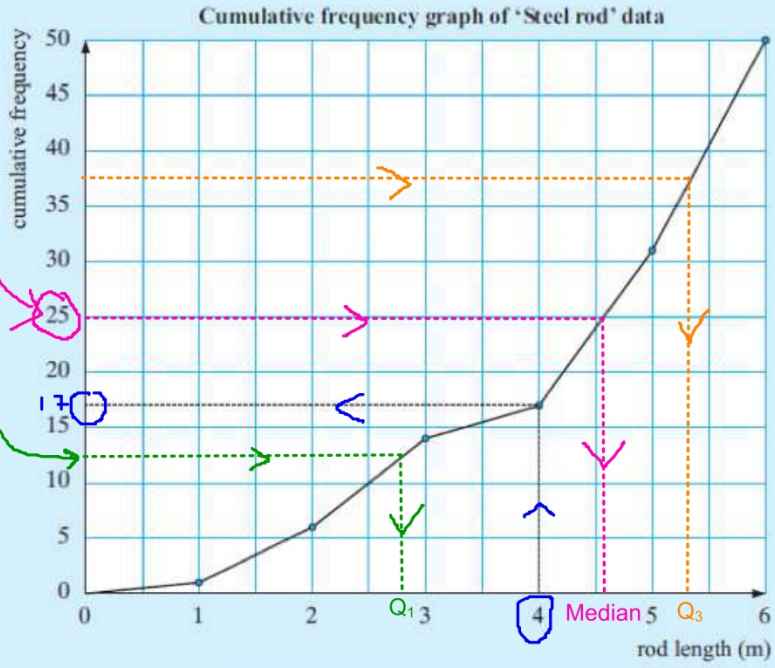
n times 25%
50 X 0.25 = 12.5

Q_1 = 2.8 ±0.05

To find the position or frequency of Q_3

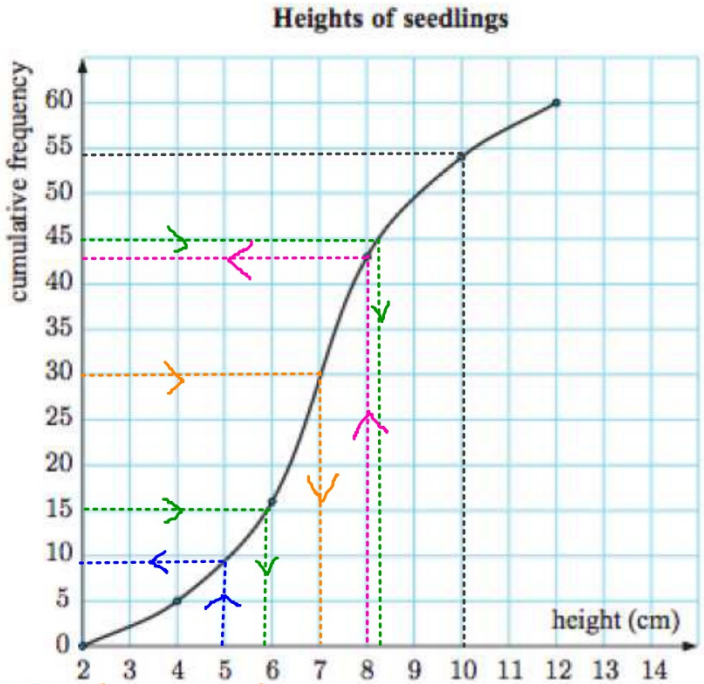
n times 75%
50 X 0.75 = 37.5

Q_3 = 5.3 ±0.05



The other way around:
How many rods measure 4 m or less: 17 rods.

4 A botanist has measured the heights of 60 seedlings and has presented her findings on the cumulative frequency graph below.



(a) How many seedlings have heights of 5 cm or less?

(b) What percentage of seedlings are taller than 8 cm?

(c) Find the median height.

(d) Find ~~the interquartile range for the heights~~ Lower & upper Quartiles

(e) Copy and complete:
"90% of the seedlings are shorter than ....." 60 x 0.9 = 54
From graph we get 10 cm

(a) Around 9

(b) Be careful. From the graph we get a frequency of 43 But that would mean, 43 seedlings are 8cm or LESS Hence the answer is 60-43=17

(c) 50% of 60 =30 frequency.
From graph we get Median = 7cm

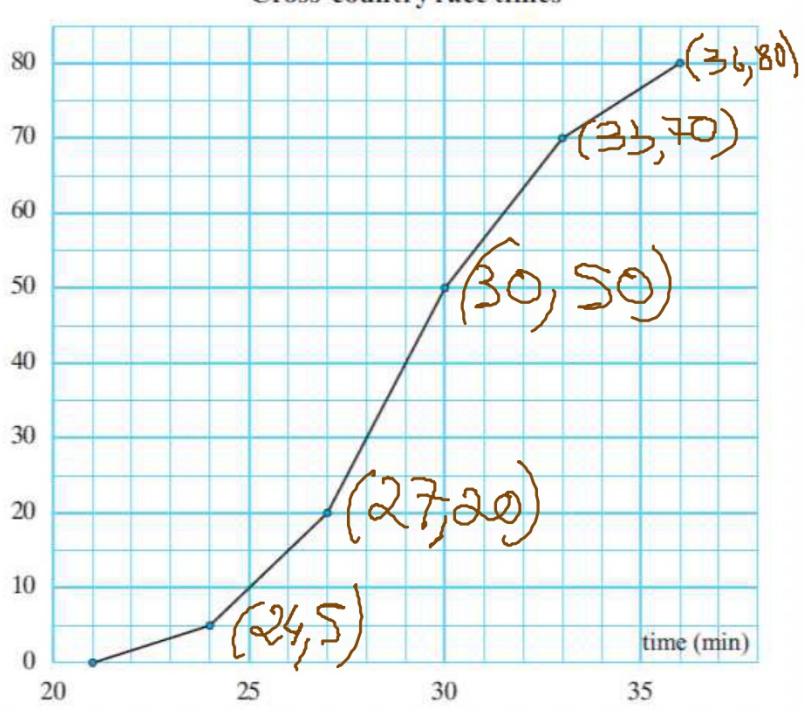
(d) Q_1 is located at a f of 15 (25 % of 60), from graph we get Q_1 = 5.85cm
Q_3 is located at a f of 45 (75% of 60), from graph we get Q_3 = 8.3cm

Can you go backwards?

From a Frequency Curve back to the Frequency Table.

Can you create Frequency Table from cumulative frequency graph?

1st get the coordinates from your points.
The "y" coordinate corresponds with the cum. *f*
The "x" coordinate will help you determine the intervals



Interval	f	Cum. f
21 ≤ t < 24	5	5
24 ≤ t < 27	15	20
27 ≤ t < 30	30	50
30 ≤ t < 33	20	70
33 ≤ t < 36	10	80
	Σ 80	

From the cumulative, you can easily work out the raw frequencies.