

Edexcel GCSE Statistics Higher between paper revision

Topics:

- Normal Distribution
- Correlation
- Petersen Capture Recapture
- Index Numbers
- Risk
- Binomial Distribution
- Box plots plus outliers
- Cumulative Frequency Graphs
- Median by interpolation
- Random Response
- Control groups
- Matched pairs
- Stem and leaf diagrams
- Choropleth Maps
- Standard deviation
- Geometric and weighted mean

Higher Tier Formulae

You must not write on this page.

Anything you write on this page will gain NO credit.

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$$

An alternative formula for standard deviation is

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$\text{Rates of change (e.g. Birth rate} = \frac{\text{number of births} \times 1000}{\text{total population}})$$

1.

For each of the following Normal distributions fill in the table for the range of values for which you find:

(a) The middle 68% of the population

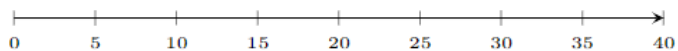
(b) The middle 95% of the population

(c) Almost all the population

Distribution	The middle 68% of the data	The middle 95% of the population	Almost all the population
$X \sim N(12, 2^2)$			
$X \sim N(17, 36)$			
$X \sim N(24.2, 0.16)$			

2. Sketch the following two normal distribution curves on the same axes, given below, showing the scale on the

	A	B
Mean	20	16
Standard Deviation	4	2



3. A company manufactures batteries with an average lifespan of 500 hours and a standard deviation of 50 hours.

a)

i. Identify the type of probability distribution that can be used to represent the lifespan of these batteries.

ii. State one condition that must be met for this distribution to be a valid model.

b) Calculate the probability that a randomly selected battery will:

i. Operate between 400 hours and 500 hours.

ii. Operate for less than 450 hours.

c) If 4000 batteries are tested, estimate how many of them would function for more than 550 hours.

4.

Below is the data for 8 runners' personal bests for a 5km race and a 10 km race

Runner	5 km time	10 km time				
A	00:18:38	00:44:30				
B	00:42:34	01:21:21				
C	00:30:54	00:54:07				
D	00:33:25	00:49:16				
E	00:22:57	00:50:01				
F	00:37:32	01:25:55				
G	00:17:05	00:47:49				
H	00:50:03	01:56:48				

(a) Calculate Spearman's rank correlation coefficient for this data.

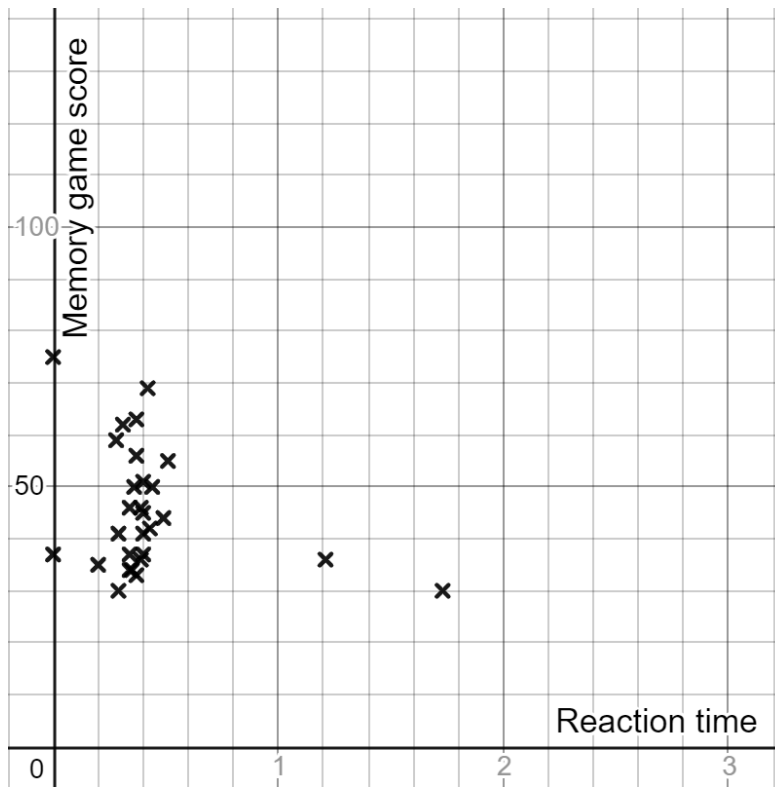
(b) Interpret this value of the Spearman's rank correlation coefficient.

(c) The 10km time for runner D has been recorded incorrectly. It should be 00:59:16

Without any further calculation, explain whether the value of the Spearman's rank correlation coefficient will be larger, smaller or stay the same.

5.

Below is the scatter graph of data showing students reaction time against their memory game score.

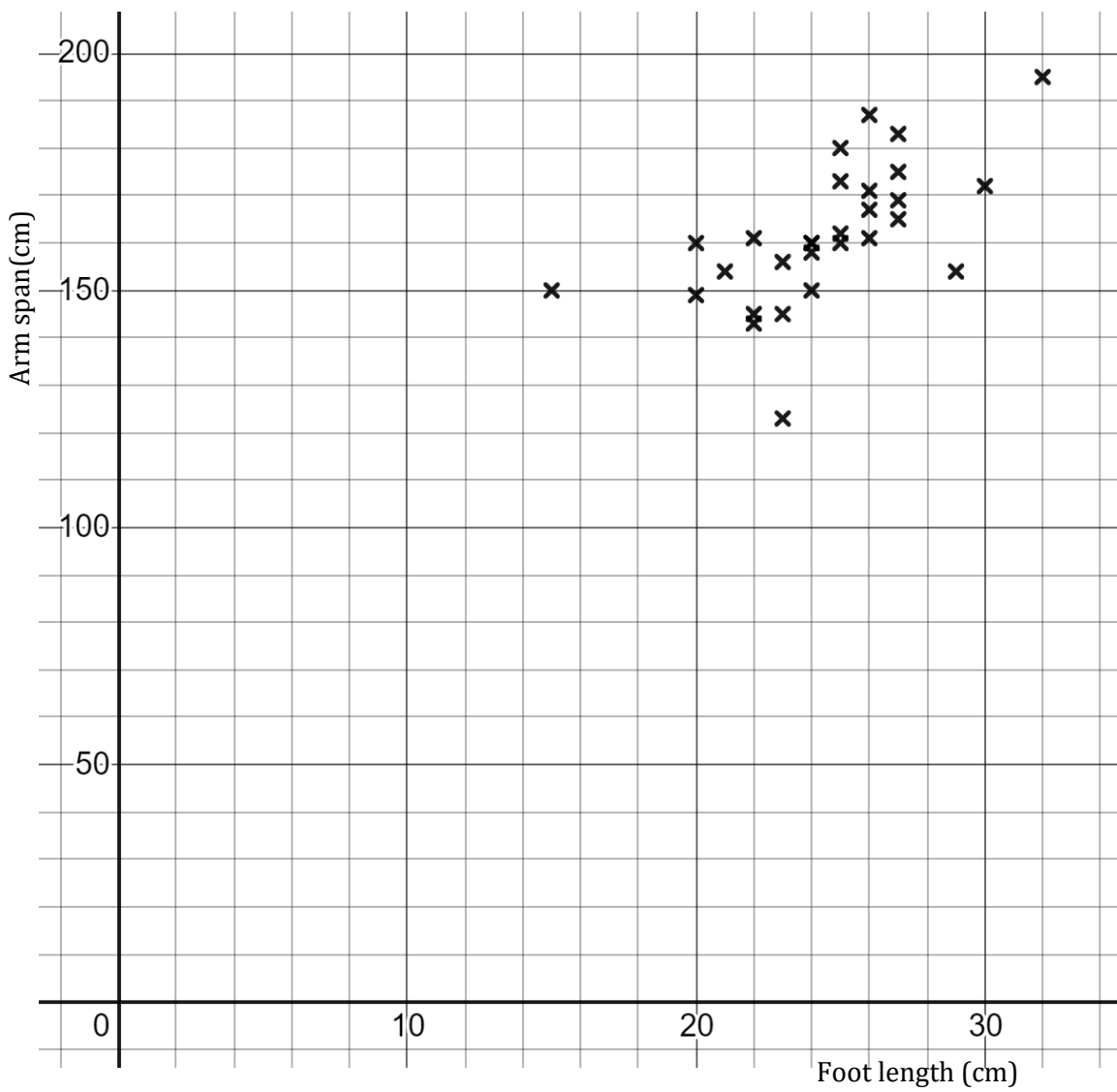


(a) Explain how you can tell this data has not been cleaned?

(b) The Pearson's correlation coefficient for this data once it has been cleaned is -0.272

Interpret this value in context.

6. The following scatter graph shows foot length against arm span.

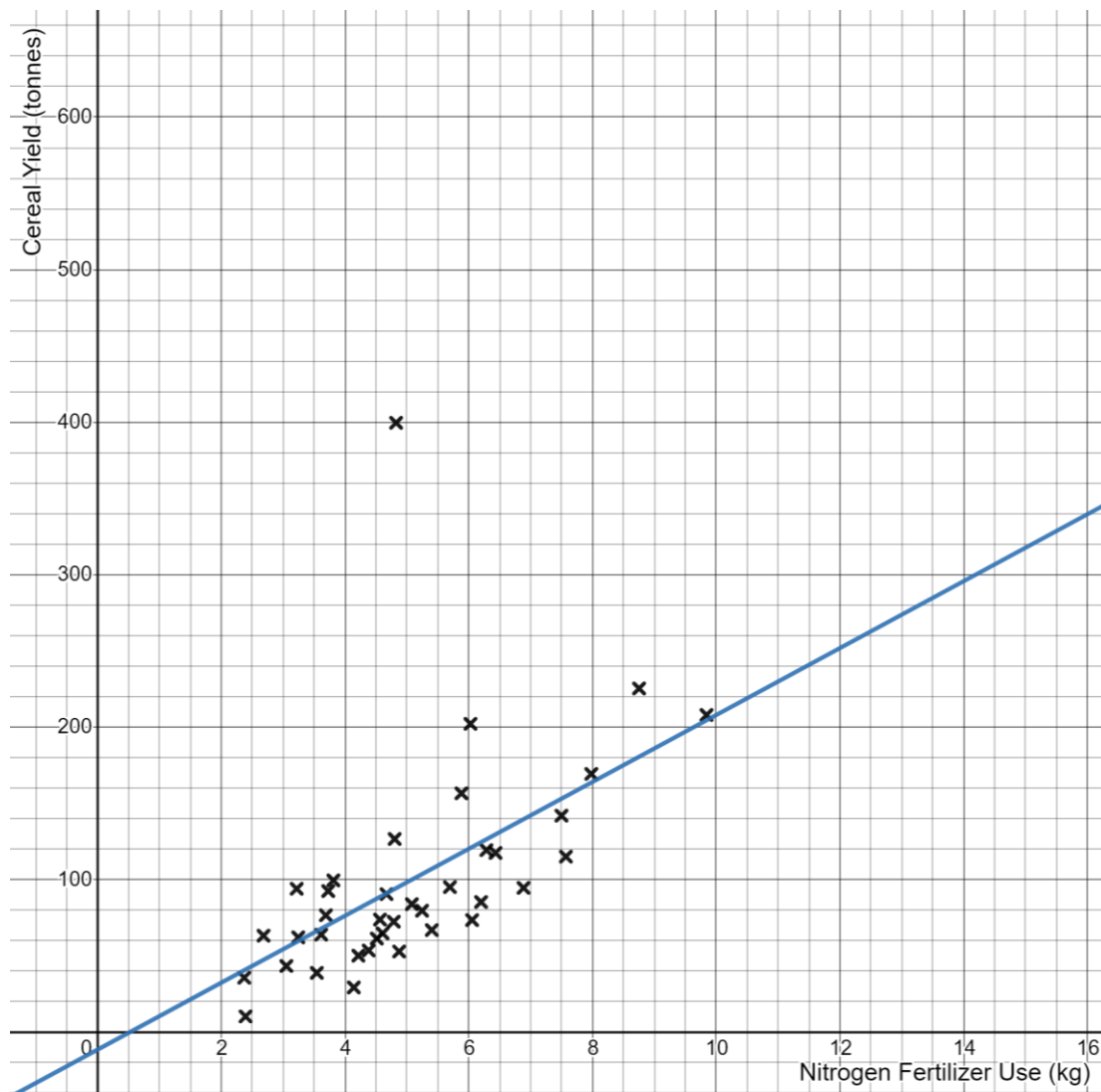


The mean point is (24.5, 161.7)

Draw a line of best fit on the scatter diagram and work out the equation of this line

7. The following scatter graph shows crop yield versus fertilizer use for the European countries in 2015.

Source: UN Food and Agriculture Organization



The equation of the line of best fit shown on the graph is

$$y = 21.9x - 11.5$$

(a) Interpret the coefficient 21.9

(b) Interpret the constant -11.5 and explain why this interpretation is unreliable.

8.

(a) What is the base year for RPI?

(b) What is the base year for CPI?

9. The table gives the monthly average price of unleaded petrol so far in 2024.

Month	Jan	Feb	Mar	Apr	May
Price per litre (pence)	139.36	141.47	144.66	149.31	
Chain base index		101.5	102.25		

(a) Calculate the chain base index numbers for April and May.
Give each value correct to 1 decimal place.

(b) Calculate the average monthly increase in petrol price per litre for 2024 so far.

10. State when this country Mexico first went into recession.
<https://data.oecd.org/chart/6EBq>



11. Below are the number of deaths for two countries.

	Country A		Country B	
Age group	Number of Deaths	Population	Number of Deaths	Population
0-29	7 000	6 000 000	6 300	1 500 000
30-59	20 000	5 500 000	3 000	550 000
60+	120 000	2 500 000	12 000	120 000

(a) Calculate the crude death rate for the 60+ age group for both countries.

(b) Given the standard population for the two countries is:

Age group	Country A	Country B
0-29	429	691
30-59	392	253
60+	179	56

Use this to calculate the standardised death rate for the 60+age group for both countries

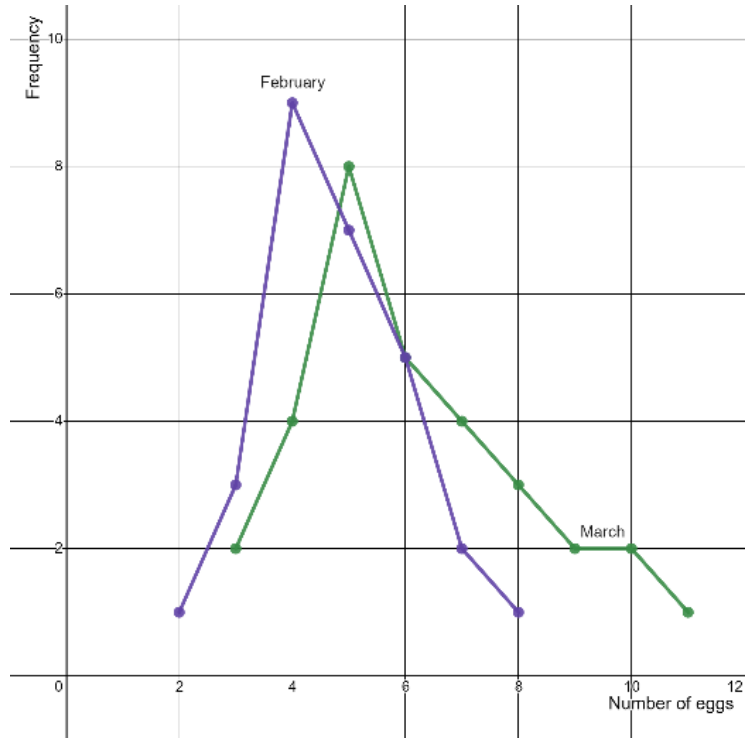
$$\text{Standardised rate} = \frac{\text{crude rate}}{1000} \times \text{standard population}$$

(c) Compare the two standardised rates

12.

The frequency polygons below show the numbers of eggs collected by Conor from his hens during February and March 2021.

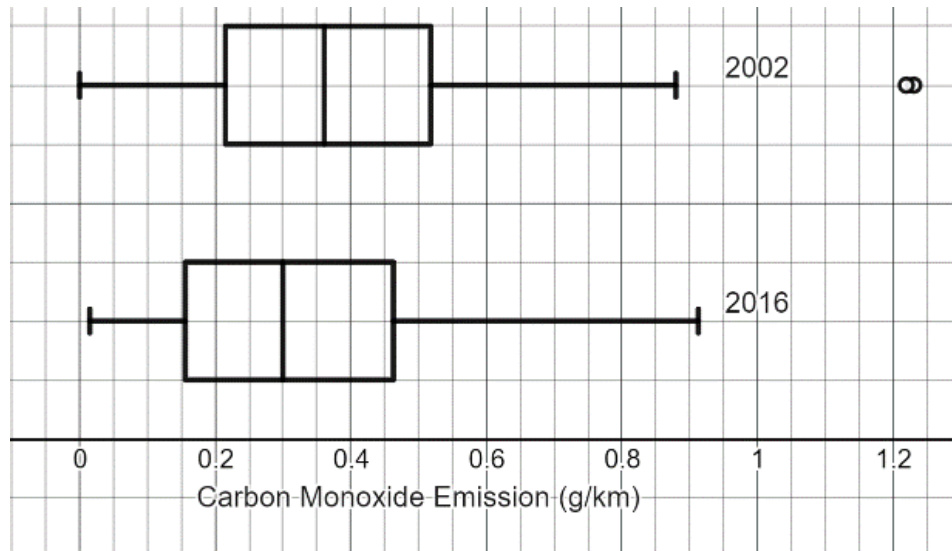
Compare the distributions



13.

The box and whisker plots below show the carbon Monoxide emission for a sample of 100 cars manufactured in 2002 and a sample of 100 cars manufactured in 2016.

Compare the distributions



14.

(a) Calculate and estimate for the mean and standard deviation of this data set, you must show all your working:

Height	Frequency					
$100 < x \leq 110$	1					
$110 < x \leq 120$	0					
$120 < x \leq 130$	1					
$130 < x \leq 140$	1					
$140 < x \leq 150$	3					
$150 < x \leq 160$	9					
$160 < x \leq 170$	9					
$170 < x \leq 180$	5					
$180 < x \leq 190$	2					

(b) Comment on the skew of the data, showing any calculations you make. (5 marks)

(c) It turns out one of the data values in the interval $180 < x \leq 190$ had been recorded incorrectly and should have been in the interval $170 < x \leq 180$.

Without any further calculations, explain the effect this would have on the mean, median and standard deviation :

Mean

Median

Standard Deviation

-
15. The percentage voter turn out for the last six general elections is given in the table below

Calculate the average voter turn out over the last six general elections

Year	Percentage turn out
2001	59.4%
2005	61.4%
2010	65.1%
2015	66.2%
2017	68.8%
2019	67.3%

16.

Tabitha collected data from her class about the number of siblings they had and their ages.

Below is an extract of her spreadsheet.

Give three reasons why she would need to clean this data:

Number of siblings	Your age	Age of sibling	Age of sibling	Age of sibling
2	13	8		
1	12	15		
0	12			
Three	12	9	14	16
1 brother, 1 sister	13	15	17	
2	1.2	5	8	
1	12			
1	13	16		

17.

Rhys has a database of information about weather in the UK at Heathrow.

Here is the information that the database contains about each day of the year in 2021.

Average temp	Max Temp	Min Temp	Max wind speed (gust)	Average wind speed	Dew point	Wind direction	Rainfall
--------------	----------	----------	-----------------------	--------------------	-----------	----------------	----------

Explain how Rhys can use technology to clean this data:

18.

Jayden wants to ask the following question in a survey:

Have you ever shoplifted? Yes No

Explain how they would use the random response technique to collect answers to the question:

Jayden thinks about 5% of people will have shoplifted at some point in their life.

Jayden gets the following data

Yes 132

No 98

Does this support his hypothesis?

20.

A butterfly farm is conducting a study to estimate the population of a particular butterfly species.

They initially captured and tagged 100 butterflies, then released them back into the farm.

After one day, they conducted a second capture and found that among the 80 butterflies captured, 20 were tagged.

Estimate the total population of butterflies in the butterfly farm.

Discuss the validity of your estimate.

21.

Below are the test results of Class A, which consisted of 31 pupils, displayed as a stem and leaf diagram.

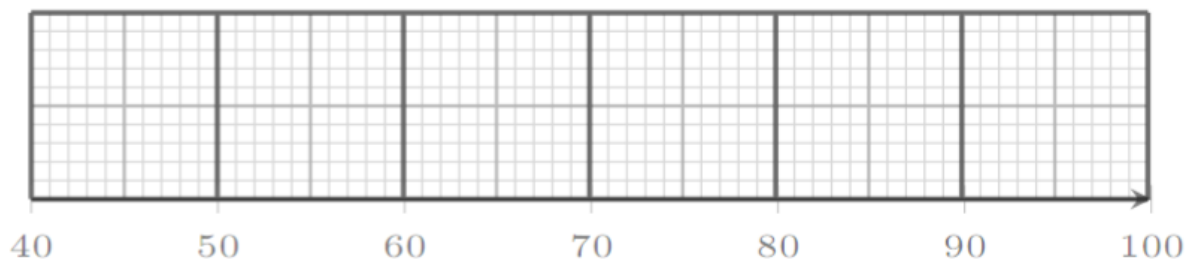
4		5
5		2
6		3
7		0 2 2 2 3 4 5 6 6 8 9
8		0 1 2 3 4 5 7 7 8 9
9		0 1 5 6 7 8 8

Key 6|3 mean 63%

(a) Calculate the value of the median, lower quartile and upper quartile

(b) Identify if there are any outliers by calculation.

(c) Draw a box and whisker plot, showing outliers if they exist.



22.

- (a) The following table gives Emily's percentage marks and their respective weights in her music exam:

	Performance	Theory	Composition
% Mark	75	80	85
Weight	40	35	25

An overall mark of 80% would give Emily a distinction for the exam. What percentage did Emily get? Did she gain a distinction?

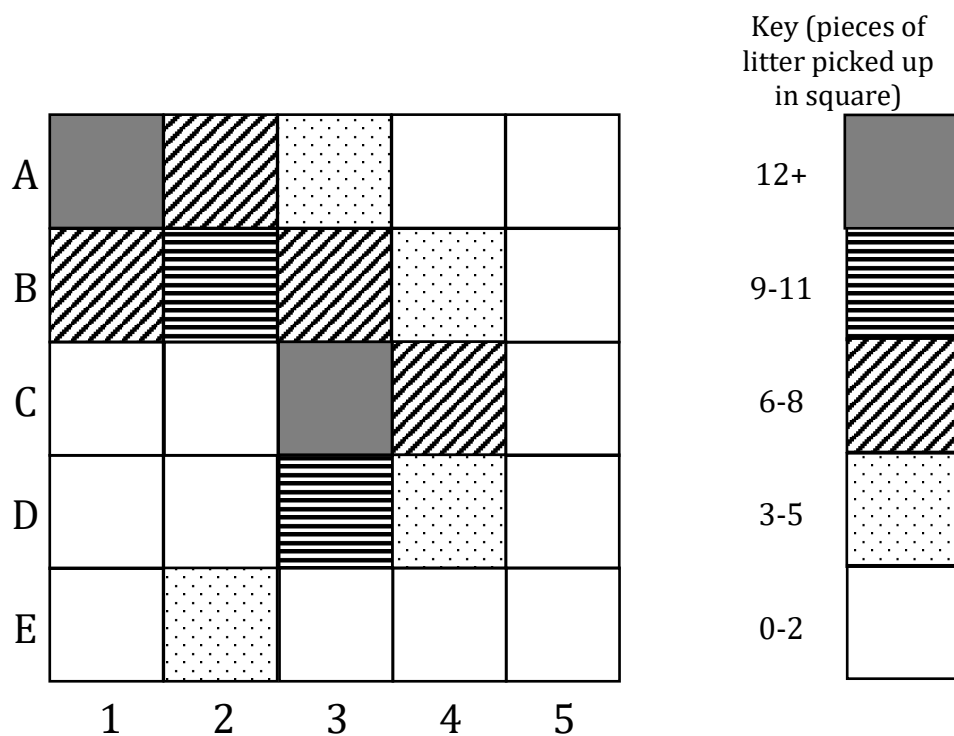
- (b) Fabio still has to complete his composition exam. What is the minimum mark he could achieve and gain a distinction?

	Performance	Theory	Composition
% Mark	80	75	
Weight	40	35	25

23. The grid below shows the amount of litter collected from the ground after lunch on a section of a school field.

A	12	8	4	0	1
B	7	9	8	3	0
C	3	6	14	6	1
D	1	5	10	4	0
E	2	3	2	0	0
	1	2	3	4	5

- (a) Complete the choropleth map by filling in squares C1, C2, D1, D2



(b) The headteacher has decided to put one more bin on this section of the field.
Explain with reasons where he should put the bin

24. Jordan has surveyed his school on their use of Vinted.

He asks people if they receive the item “as described” or “not as described” and what the average five star rating of the seller is.

He displays his results in a table.

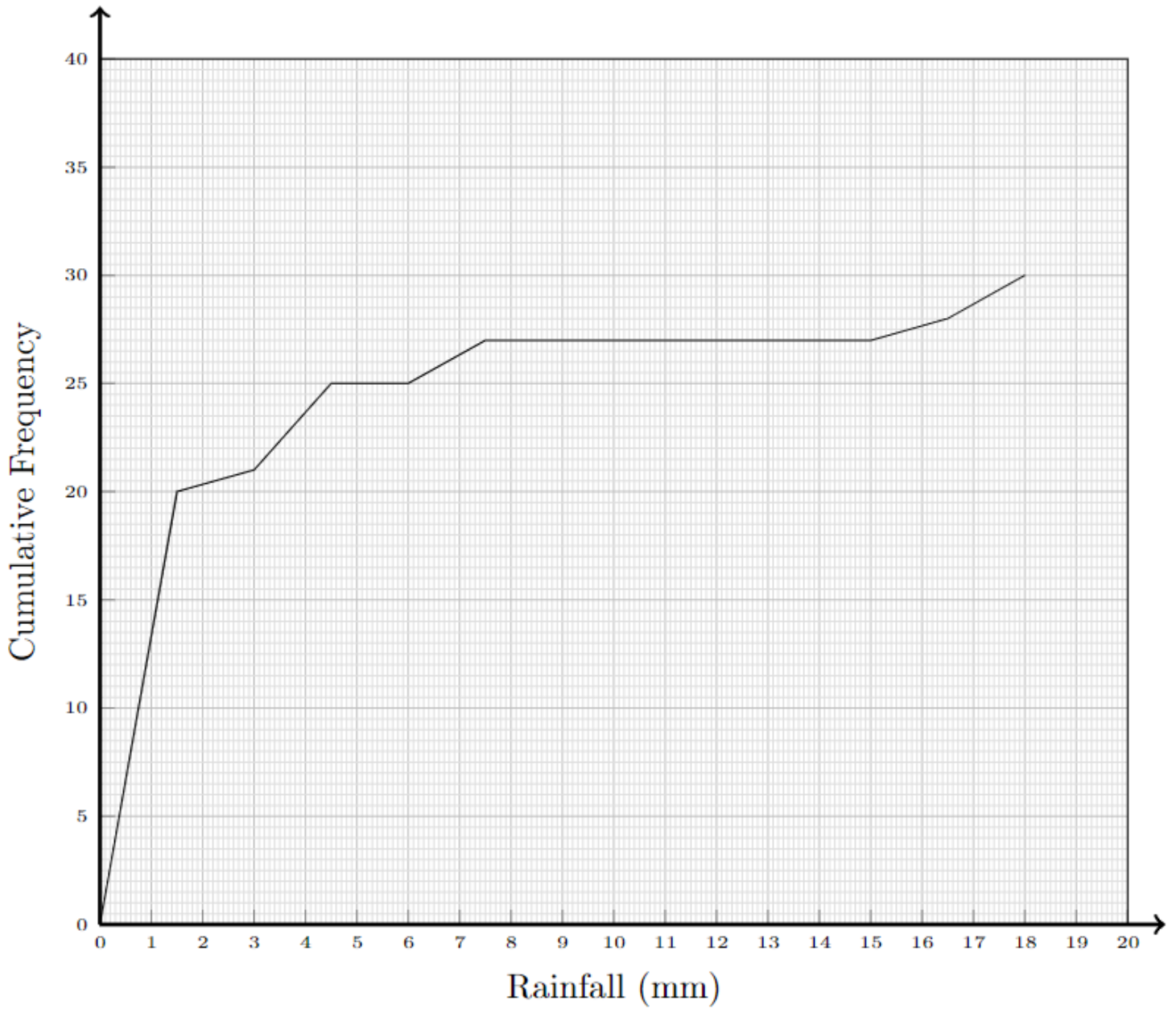
	As described	Not as described
4 stars or more	80	3
Less than 4 stars	45	10

Jordan says the relative risk of receiving an item “not as described” when the seller has an average rating of less than 4 stars, compared to when the seller has an average rating of 4 stars or more, is approximately 5.

(a) Show that Jordan is correct

(b) Interpret a relative risk of 5 in context.

25. The cumulative frequency graph shows the daily rainfall in Reigate during April 2024



(a) Find the 10th to 90th interpercentile range for rainfall in Reigate in April 2024

26. On a fruit stall at the market, the probability of an apple being bruised is 0.08.

Bradley buys 9 apples.

(a) Find the probability that 6 apples are bruised.

(b) Discuss the validity of the assumptions you have made.