Edexcel Higher Stats GCSE 2025
Paper 2 possible topics

Possible topics (high)

- Describing populations/samples/sampling frames
- Random number generation/use of tables
- Bias in non-random sampling
- Stratified sampling
- Systematic sampling
- Cleaning data
- Composite bar charts
- Venn diagrams
- Comparative pie charts
- Geometric mean
- Weighted mean
- Transforming averages
- Standard deviation calculations
- Box plots (including outliers)
- Cumulative frequency
- Scatter diagrams
- Regression lines
- Index numbers
- Risk
- Normal distribution
- Quality assurance

Possible topics (low)

- Median by interpolation
- Time series (from graph)
- Binomial distribution including conditions

1) Starting from Row 1 Column 1 and moving horizontal generate 8 random numbers in the range 00-59.
Show all your working.

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

~~62~~ 04 13 ~~77~~ 33 46

44 ~~73~~ ~~78~~ ~~94~~ 18, 07, 10

04, 13, 33, 46, 44, 18, 07, 10

2) A youth club of 67 members wants to find out what its members think about the summer holiday activities it offers.
The club has a full list of members in age order and decides to use **systematic sampling**.
They select every 5th person on the list to complete a short survey.

Discuss the advantages and disadvantages of this sampling method.

BIAS, COST, SAMPLE SIZE, TIME

ADV:

Bias: Because the list is in age order this should produce an age representative sample.

(Although not everyone has an equal chance of being selected. The final two people on the list may never get selected.)

ADV:

Cost/Time: This is a relatively cheap and quicle method for sampling the members.

DIS:

SAMPLE SIZE: This produces a sample size of B which is quite small.

3) A school wants to carry out a survey to find out students' opinions on school lunches. The school has 600 students in total. The breakdown of students by year group and gender is shown in the table below:

Year Group	Number of Boys	Number of Girls	Total
Year 9	120	80	200
Year 10	90	60	150
Year 11	60	40	100
Year 12	70	80	150
Total	340	260	600

Liam decides to survey 120 students.

(a) Liam calculates that he needs to survey 24 year 9 boys. Show that he is correct.

$\frac{120}{600} \times 120 = 24$

(b) Liam gave his sample a form to complete. This is a section of the spreadsheet that resulted.

Year Group	Lunch Type	Quality Rating	Likes Most	Healthy Eating Importance	Variety Wanted
Yr9	school	good	hot meals	very important	YES
10	packed	n/a	-	somewhat	no
year11	School Lunch	avg	convienience	not very	yes
Y12	school lunch	poor	cheap	v. important	y
9	packed lunch			very	?
Year 19	school	Excellent	tasty	some	yes
11	school lunch	bad	nothing	very important	YES
yr12	packed	n/a	n/a	not at all	no
Y9	School Lunch	Good	friendly staff	somewhat important	Yes
10	school lunch	ok	quick service	very important	yes

(i) Explain how Liam should deal with the "Year 19" entry in the first column.

- This is likely a typo. He should correct it to 10.

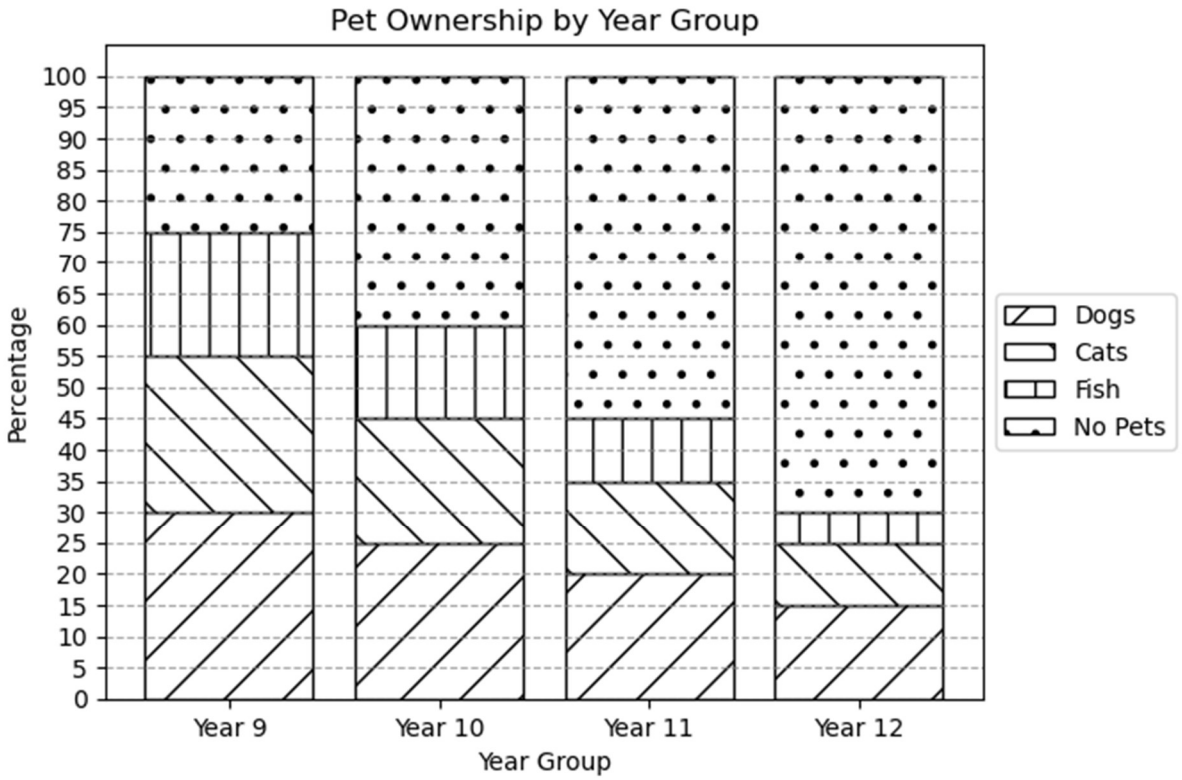
or He should delete the entire row.

(ii) Explain how Liam could use technology to clean the data in the "Lunch Type" column.

Liam could use the find and replace function to find and replace all instances of the word "lunch" with an empty character.

4) Below is a composite bar chart of pet ownership by year group for the school.

Year Group	Total
Year 9	200
Year 10	150
Year 11	100
Year 12	150
Total	600



(a) Calculate the total number of students who don't own a pet.

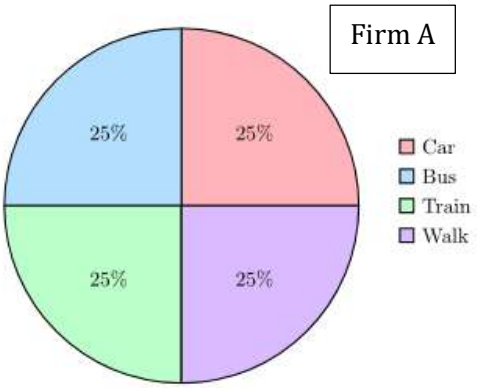
$200 \times 0.25 + 150 \times 0.4 + 100 \times 0.55 + 150 \times 0.7$

$= 270$

5) Rio has collected data on commuter methods for two different firms in his town.

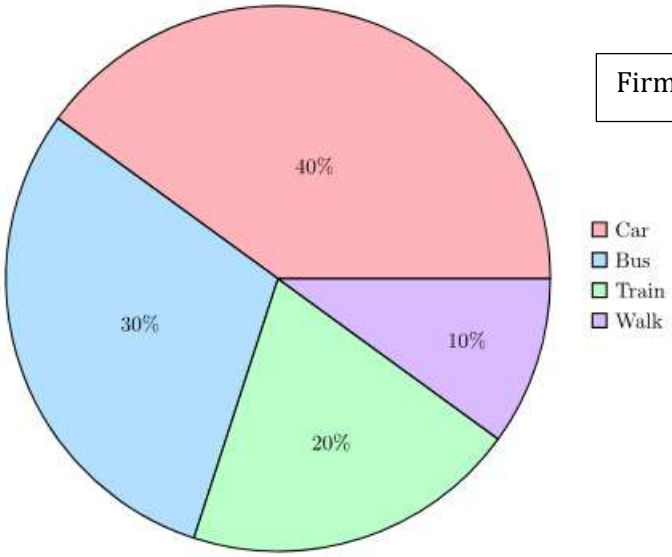
He has drawn the comparative pie charts shown below.

Firm A



NOT DRAWN TO SCALE

Firm B



(a) Give a reason why Rio would have drawn comparative pie charts.

He collected data from a different number of people at each firm

(b) 32 people commute by car in Firm B. Find out the number of people Rio asked in Firm B.

32	40
80	100

×2.5

people.

(c) In his final pie charts the radius for firm A was 3cm and the radius for firm B was 4.5cm.
Calculate how many people commute by car in the firm A sample.

80	180
$\pi(3^2)$	$\pi(4.5)^2$

$\times \frac{9}{4}$

180 people

45 people by car

6) Below is the average price of hard cheese per kg for 2015 – 2024

Year	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Price per kg	766.6	718.8	718.2	727.7	708.7	688.5	628.3	709.1	918.5	837.6
Simple Index Number	100	93.8	93.7	94.9	92.4	89.8	82.0	92.5	119.8	109.3

(a) Taking 2015 as the base year, calculate the simple index number for 2018.

$$\frac{727.7}{766.6} \times 100 = 94.9$$

(b)

The CPI compared to 2015 is shown below as simple index numbers

Compare the index numbers for cheese to the CPI.

Year	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
CPI	100.0	100.7	103.4	105.3	104.3	102.7	103.5	111.9	117.1	110.0

For 2015 – 2022 the simple index number for cheese was below 100, whereas the CPI was above 100.

In 2023 & 2024 the index numbers for cheese were higher than CPI.

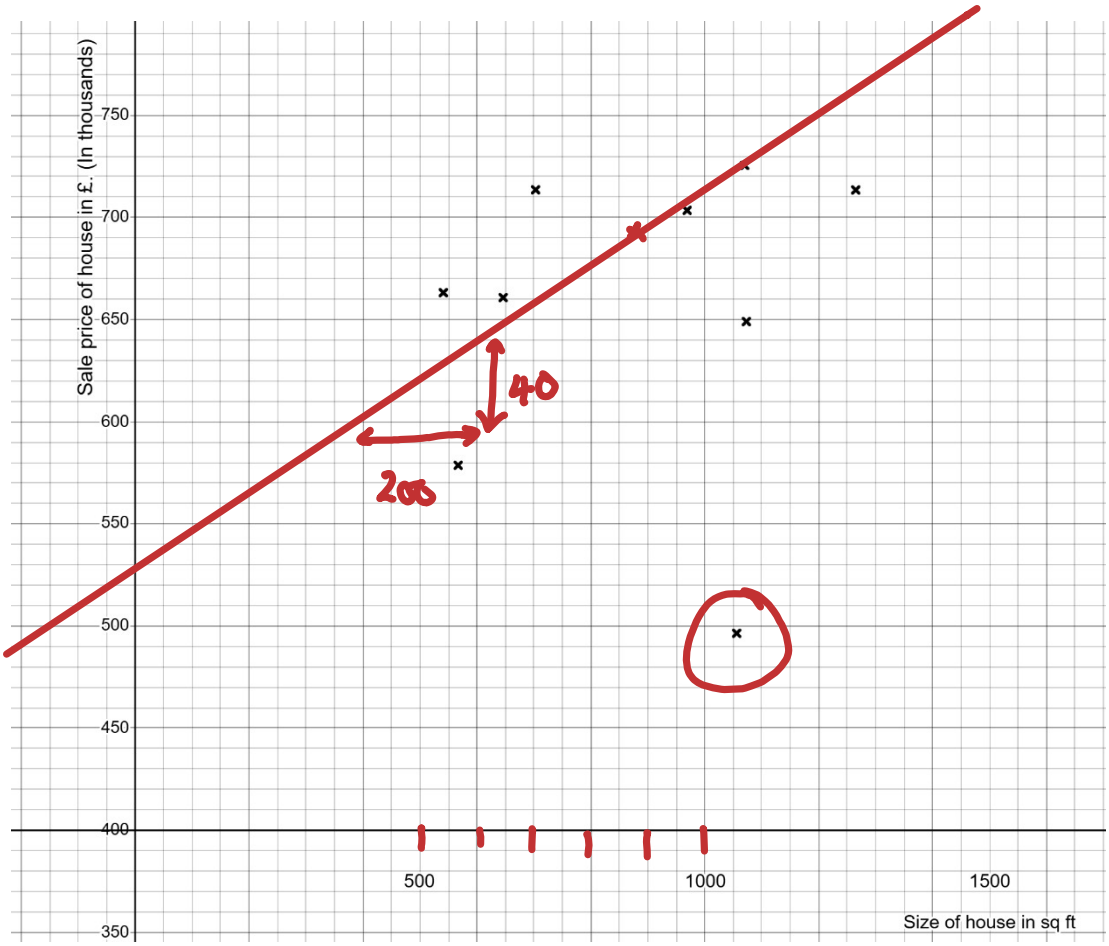
7) A housing analyst is comparing how house prices relate to property size.

Below is the data for several London houses, based on recent sales.

Size (sq ft)	Price
~~1056~~	~~496320~~
1265	713460
541	663266
703	713545
1147	848780
1073	649165
969	703494
646	660858
567	578907
1070	725460

887, 695215

These have been plotted on the scatter diagram below.



(a) Identify an outlier from the scatter diagram and ring this value on the scatter diagram.

(b) Ignoring the outlier calculate the mean point $(\bar{x}, \bar{y})$ and plot this on the scatter diagram.

$\bar{x} = \frac{7981}{9} = 887$ $\quad \bar{y} = \frac{6256935}{9} = 695215$

(c) Draw a linear regression line for the data, ignoring the outlier.

(d) Calculate the regression line equation for London

$y = 200x + 525000$ (correct eqn for your line)

Liam also calculates regression line equations for Manchester, Birmingham, Galsgow and Cardiff.
These are listed in the table below.

x=size of property in square foot
y=sale price of property in £.

City	Regression Line
Manchester	$y = 140x + 380000$
Birmingham	$y = 130x + 365000$
Glasgow	$y = 120x + 270000$
Cardiff	$y = 135x + 372000$

(e) Compare the regression lines for Birmingham and Cardiff in context.

- Birmingham and Cardiff lines both have a positive gradient show that as the property size increases so does the price.
- Cardiff has a larger gradient by 5 show that each square foot costs approximately £5 in Cardiff.
- Cardiff has a higher y-intercept by £7,000 showing that the cost of fees in Cardiff is slightly higher.

8) A recent study looked at the effect of a new energy drink on the likelihood of developing frequent headaches.

2226 people were surveyed.

The results are displayed in the table below.

	Frequent Headaches	No Frequent Headaches	Total
Drinkers	97	1,040	1,137
Non-Drinkers	46	1,043	1,089
Total	143	2,083	2,226

(a) Calculate the absolute risk of developing frequent headaches after consuming the energy drink.

$\frac{97}{2226} = 0.044$

(b) Calculate the relative risk of developing frequent headaches after consuming the energy drink compared to not consuming the energy drink.

(c) Comment on this risk value in context.

b) $\dfrac{\frac{97}{1137}}{\frac{46}{1089}} = \dfrac{0.085}{0.042} = 2.02$

c) You are twice as likely to develop frequent headaches after drinking the energy drink, than if you don't drink it.

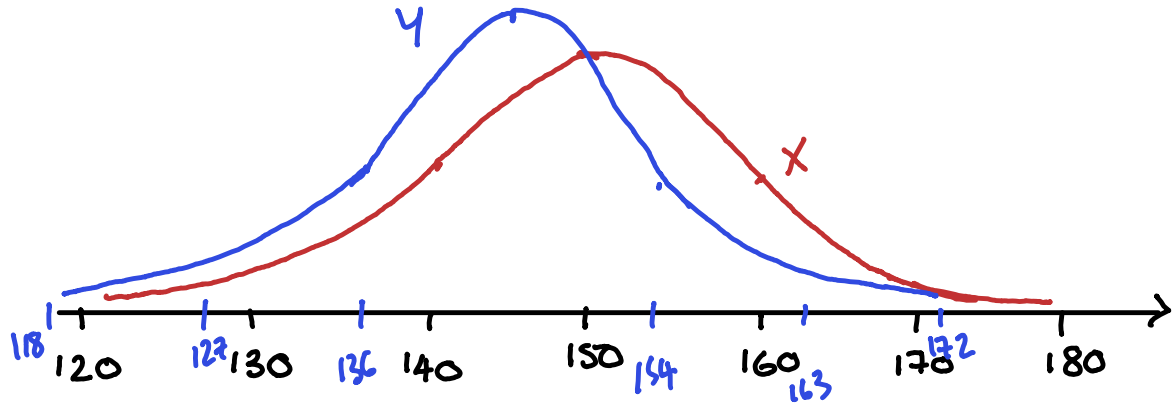
9)

The weights of apples from two different orchards are normally distributed:

Orchard X: $N(150,100)$

Orchard Y: $N(145,81)$

On the same set of axes, sketch the two normal distribution curves. Clearly indicate the mean and spread of each.



Which orchard produces apples with more consistent weights? Justify your answer.

Y because it has a smaller standard deviation.

10) A teacher records the heights of 200 students and finds that the data is approximately normally distributed with a mean of 170 cm and a standard deviation of 6 cm.

Estimate how many students are expected to have heights between:

- (a) 164 cm and 176 cm
- (b) 158 cm and 182 cm

$X \sim N(170, 6^2)$

$164 \leq x \leq 176$ $\mu \pm \sigma$ 68%

$0.68 \times 200 = 136$

$158 \leq x \leq 182$ $\mu \pm 2\sigma$

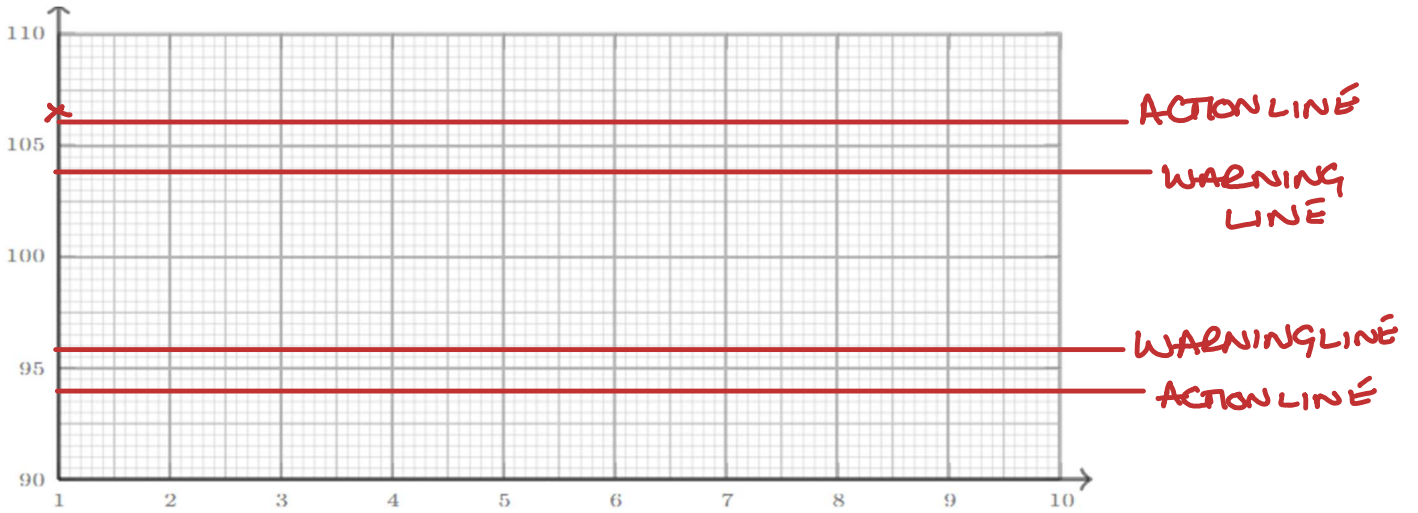
$0.95 \times 200 = 190$

11) A factory produces metal rods. The target length is 100 mm with a standard deviation of 2 mm. The factory uses control charts to monitor the process.

$100-4=96$ $100-6=94$

$100+4=104$ $100+6=96$

(a) Complete the control below.



(b) The first sample taken is 106.5mm. Mark this on your control chart and identify what action the factory should take.

Stop the machinery and reset it.

12) The table shows the time (in minutes) that a sample of students spent on a homework task

Time (minutes)	Frequency		
$0 < t \leq 10$	4	× 5	= 20
$10 < t \leq 20$	7	× 15	= 105
$20 < t \leq 30$	12	× 25	= 300
$30 < t \leq 40$	6	× 35	= 210
$40 < t \leq 50$	1	× 45	= 45
	30		

(a) Calculate the mean time spent on homework by these students.

$\frac{680}{30} = 22.67$

(b) Given that $\Sigma fx^2 = 18550$ calculate the standard deviation, showing all your working.

$\sigma = \sqrt{\frac{18550}{30} - 22.67^2} = \sqrt{\frac{941}{9}} = 10.23$

(c) A second class has a mean of 19.2 and a standard deviation of 6.2 minutes.
Compare and comment on these distributions.

The first class has a higher mean by 3.47.

The first class has a higher standard deviation by 4.03.

The first class on average spent longer on the homework task, but their times were more variable.

13) A Business studies examination has three components.

The final mark is calculated using the following weighting:

Component	Weighting
Coursework	40
Written Exam	50
Presentation Task	10

Jenna scored 72% on the coursework and 80% on the presentation task.

To achieve a distinction she needs an average of 75%.

Work out the minimum mark she can achieve on the written exam and still be awarded a distinction.

$$\frac{40 \times 72 + 80 \times 10 + x \times 50}{100} = 75$$

$$50x = 7500 - 2880 - 800$$

$$50x = 3820$$

$$x = 76.4\%$$

14) A savings account advert claims:

"**Earn an average of 5% interest per year over 5 years!**"

The actual interest rates for the account over the 5 years were:

Year	Interest Rate (%)
1	2.0
2	3.5
3	6.0
4	7.5
5	6.0

Comment on whether the advert is misleading. Justify your answer using your calculations

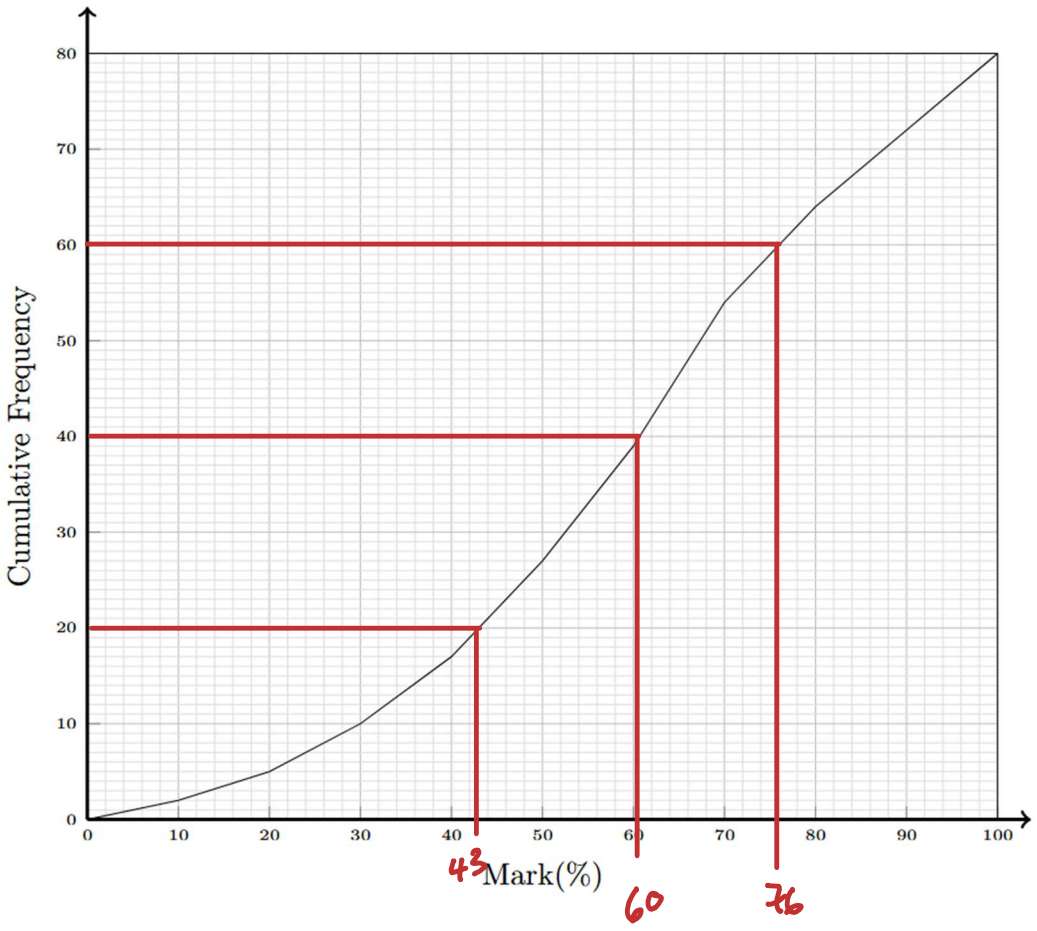
$$\sqrt[5]{1.02 \times 1.035 \times 1.06 \times 1.075 \times 1.06}$$

$$= 1.0498$$

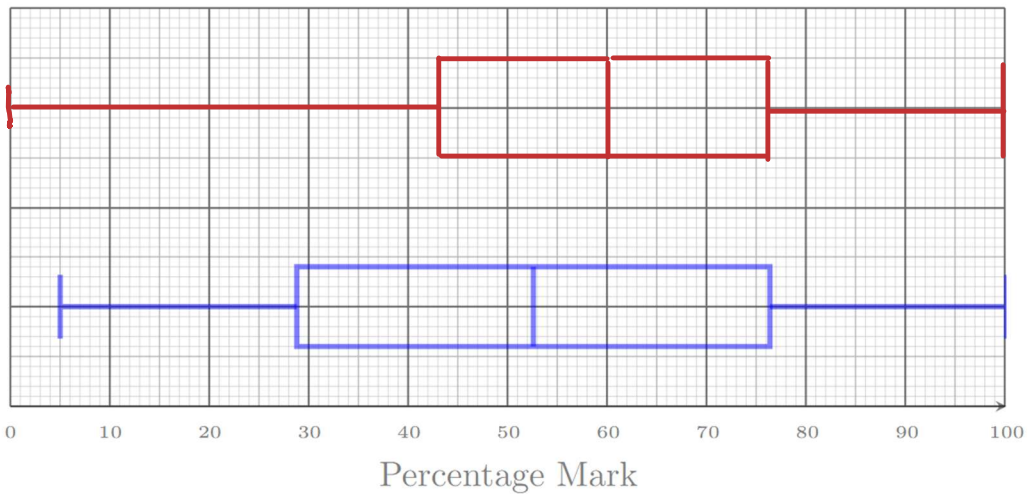
The geometric mean for these five years is 1.0498, showing an average interest rate of 4.98%, so the advert is not misleading

15) A school arere comparing results for two classes of students.

The teacher from class A has draw a cumulative frequency graph as shown below:



The teacher from class B has drawn a boxplot as shown below:



(a) Draw the boxplot for class A on the same axes as the boxplot for class B

(b) Compare the distribution of marks for the two classes.

Class A has a large median than Class B by 8.

Class A has a smaller IQR of 34, compared to class B of 48

Both classes have a symmetrical distribution.

16) In a group of 100 students:

52 study Mandarin (M)

48 study Spanish (S)

40 study French (F)

20 study both Mandarin and Spanish

18 study both Mandarin and French

15 study both Spanish and French

10 students study none of the three subjects

F: $40-18+x-15+x-x$

$= 7+x$

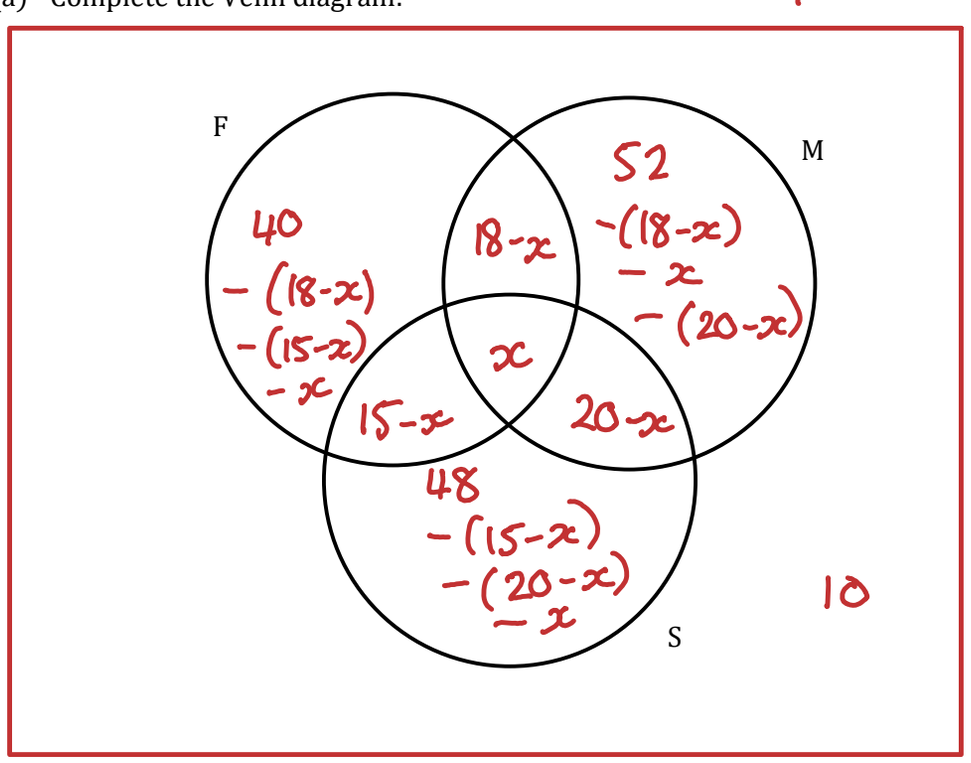
S: $48-15+x-20+x-x$

$= 13+x$

M: $52-18+x-x-20+x$

$= 14+x$

(a) Complete the Venn diagram:



$7+x+13+x+14+x+18-x+15-x+20-x+x+10=100$

(b) Given that a student studies French, calculate the probability that they also study Spanish.

b)

$\frac{15}{40} = \frac{3}{8}$

$97+x=100$

$x=3$