Q1

Ten candidates were ranked as follows by two independent examiners, according to the score they obtained in an interview. [4]

Candidate Number	1	2	3	4	5	6	7	8	9	10
Ranked by Ex. 1	7	9	1	3	8	4	10	5	6	2
Ranked by Ex. 2	9	5	1	4	6	7	8	2	10	3

Calculate the Spearman's rank correlation Coefficient and interpret the result.

Q2

A department store has the following statistics of sales (Y) for a period of 2 years of 10 salespersons who have varying years of experience (X) in sales promotion.

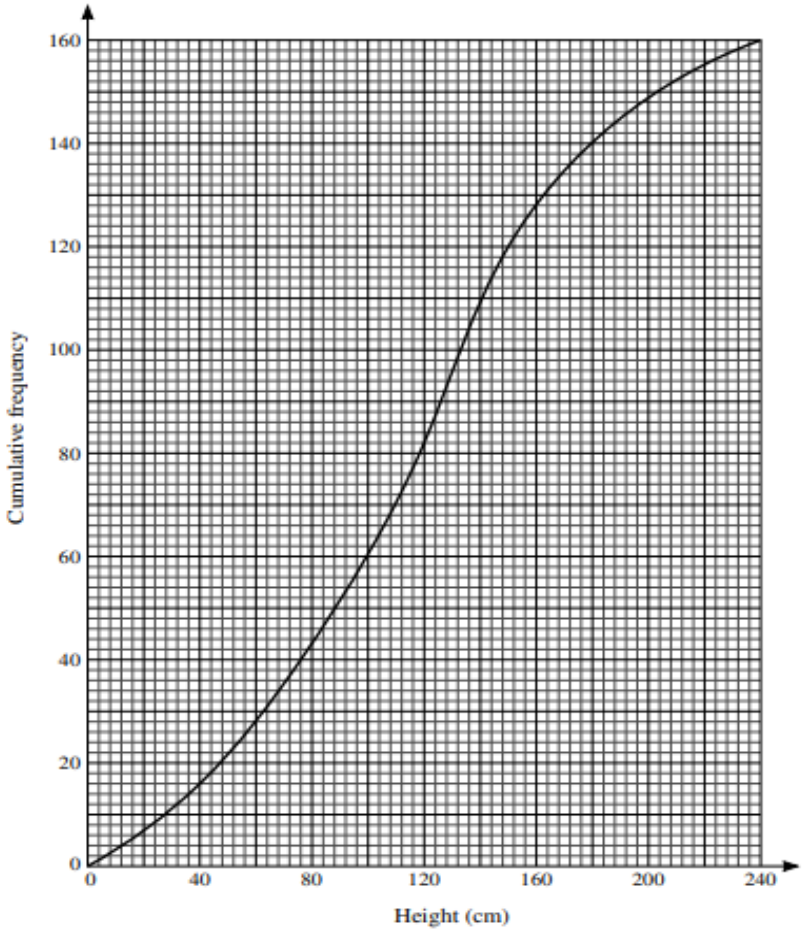
Salesperson	1	2	3	4	5	6	7	8	9	10
Experience (X) in Years	1	3	4	4	6	8	10	10	11	13
Average Annual sales (Y) in thousand	80	97	92	102	103	111	119	123	117	136

(a) Find the regression line of Y on X in the form $Y = a + bX$.

(b) Predict the annual sales volume of persons what have 12 and 15 years' experience.

Q3

The heights in cm of 160 sunflower plants were measured. The results are summarised on the following cumulative frequency curve.



(a) Use the graph to estimate the number of plants with heights less than 100 cm.

(b) Use the graph to estimate the 65th percentile of the distribution.

(c) Use the graph to estimate the interquartile range of the heights of these plants.

Q4.

The heights, in cm, of the 11 basketball players in each of two clubs, the Amazons and the Giants, are shown below.

Amazons	205	198	181	182	190	215	201	178	202	196	184
Giants	175	182	184	187	189	192	193	195	195	195	204

By using Box and Whisker plot illustrate and compare the above information.

Find the interquartile range of the heights of the players in the Amazons.

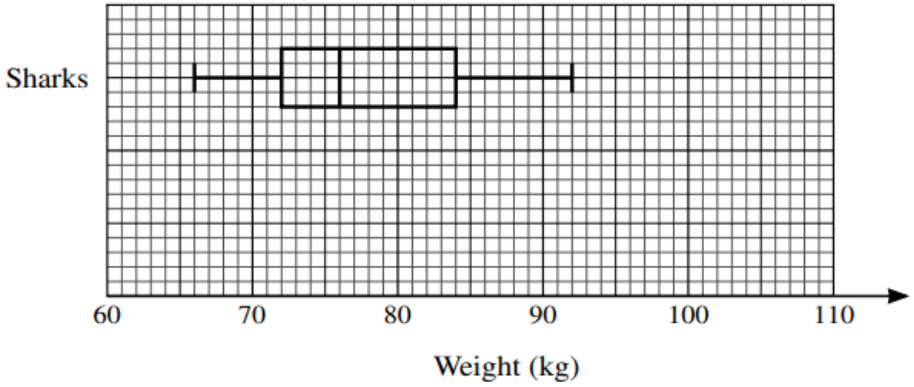
Q5.

The weights, in kg, of 15 rugby players in the Rebels club and 15 soccer players in the Sharks club are shown below.

Rebels	75	78	79	80	82	82	83	84	85	86	89	93	95	99	102
Sharks	66	68	71	72	74	75	75	76	78	83	83	84	85	86	92

Find the median and the interquartile range for the Rebels.

A box-and-whisker plot for the Sharks is shown below.



(c) On the same diagram, draw a box-and-whisker plot for the Rebels. [2]

(d) Make one comparison between the weights of the players in the Rebels club and the weights of the players in the Sharks club. [1]

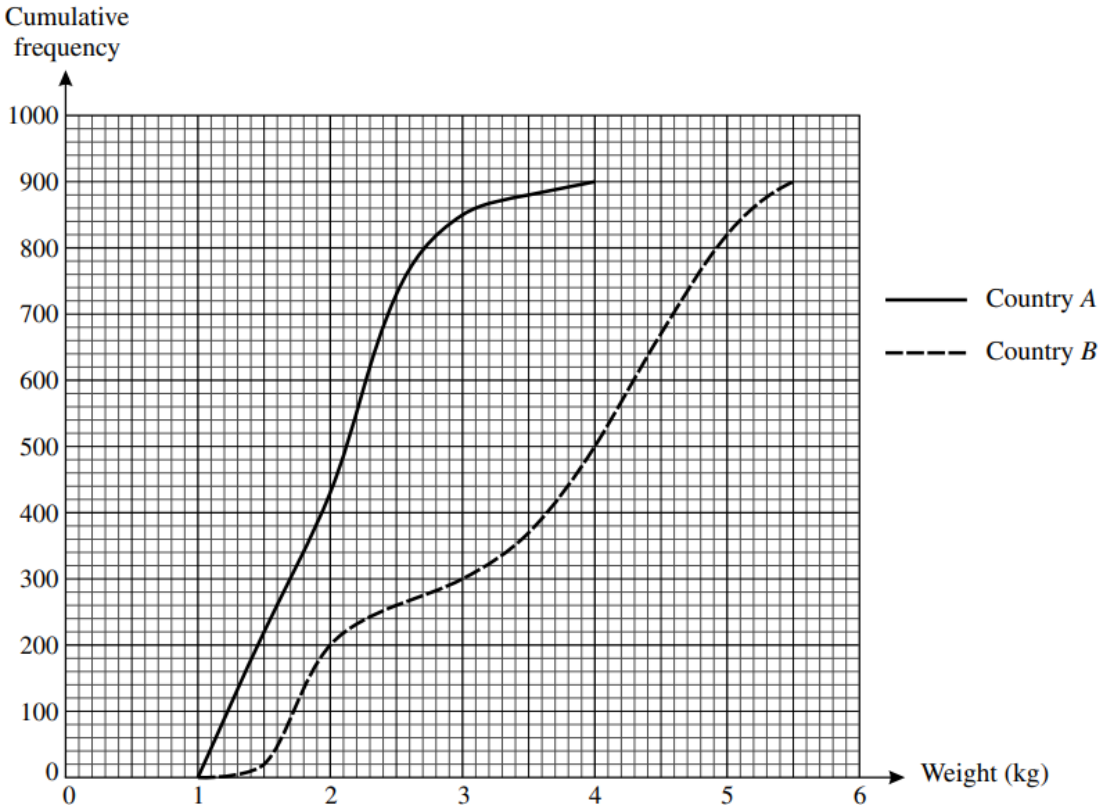
Q6 (a)

The following table summarizes the performance analysis of three teams in a football tournament.

Interpret the result mentioning which team's performance is better and consistent.

	Mean Number Goal	Standard deviation
Team A	3.15	1.37
Team B	2.9	0.29
Team C	1.23	2.05

(b)



The birth weights of random samples of 900 babies born in country A and 900 babies born in country B are illustrated in the cumulative frequency graphs. Use suitable data from these graphs to compare the central tendency and spread of the birth weights of the two sets of babies.

Q7

(a) The following back -to-back stem and leaf diagram shows the cholesterol count for a group of 45 people who exercise daily and for another group of 63 who do not exercise. The figures in brackets show the number of people corresponding to each set of leaves. [6]

	People who exercise		People who do not exercise	
(9)	9 8 7 6 4 3 2 2 1	3	1 5 7 7	(4)
(12)	9 8 8 8 7 6 6 5 3 3 2 2	4	2 3 4 4 5 8	(6)
(9)	8 7 7 7 6 5 3 3 1	5	1 2 2 2 3 4 4 5 6 7 8 8 9	(13)
(7)	6 6 6 6 4 3 2	6	1 2 3 3 3 4 5 5 5 7 7 8 9 9	(14)
(3)	8 4 1	7	2 4 5 5 6 6 7 8 8	(9)
(4)	9 5 5 2	8	1 3 3 4 6 7 9 9 9	(9)
(1)	4	9	1 4 5 5 8	(5)
(0)		10	3 3 6	(3)

Key: 2 \| 8 \| 1 represents a cholesterol count of 8.2 in the group who exercise and 8.1 in the group who do not exercise.

(i) Give one useful feature of a stem and leaf diagram

(ii) Find the median and the IQR of the cholesterol count for the group who do not exercise.

You are given that the lower quartile, median and upper quartile of the cholesterol count for the group who exercise are 4.25, 5.3 and 6.6 respectively.

(iii) On a single diagram on graph paper, draw two box and whisker plots to illustrate the data.

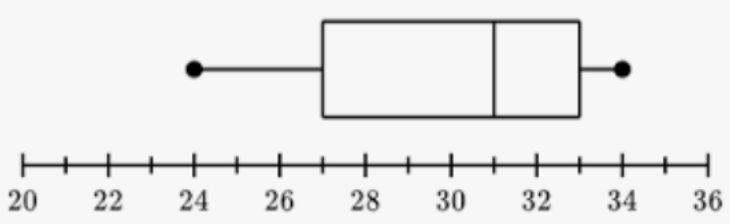
(b) During January the numbers of people entering a store during the first hour after opening were as follows [4]

Time after opening, x minutes	Frequency	Cumulative frequency
$0 < x \leqslant 10$	210	210
$10 < x \leqslant 20$	134	344
$20 < x \leqslant 30$	78	422
$30 < x \leqslant 40$	72	a
$40 < x \leqslant 60$	b	540

(i) Find the values of a and b.
(ii) Find the mean entering time

Q8

Find the range and the interquartile range of the dataset represented in the following box and whisker plot. What type of skewness would you expect this set of data to have?



Q9

The following figures are the amount spent on food by a family for 11 weeks

$ 16 $12 $9 $11 $7 $19 $15 $12 $10 $20 $13

(a) Find standard deviation
(b) Justify whether this set of data has any outlier or not?

Q10

The back-to-back stem-and-leaf diagram shows the diameters, in cm, of 19 cylindrical pipes produced by each of two companies, A and B.

Company A							Company B				
					4	33	1	2	8		
	9	8	3	2	0	34	1	6	8	9	9
8	7	5	4	1	1	35	1	2	2	3	
		9	6	5	2	36	5	6			
			4	3	1	37	0	3	4		
						38	2	8			

Key: 1 | 35 | 3 means the pipe diameter from company A is 0.351 cm and from company B is 0.353 cm.

(i) Find the interquartile rang of the diameter of the pipes produced by the companies A and B.
(ii) Comment on the nature of distribution for each set of data.

Q11

The summary of a set of 15 data is given as follows: $\sum x^2 = 240$, $\sum x = 45$. Calculate the standard deviation and hence find the value of $\sum(x - \bar{x})^2$.

Q12

The following table summarizes the performance analysis of two teams in a football tournament. Interpret the result.

	Mean Number Goal	Standard deviation
Team A	3.15	1.37
Team B	2.9	0.29

Q13

The following back-to-back stem and leaf diagram represents the monthly salaries, in dollars, of 27 employees at each of two companies, A and B.

Company *A*		Company *B*
5 4 1 1 0	25	4 4 5 6 6 7
9 9 8 7 2 1 0	26	0 1 3 5 5 7 9 9
8 6 4 2 1 0	27	1 3 4 6 6 8 8
6 5 4 2 0	28	0 1 2 2 2
9 8 5	29	
1	30	9

Key: 1 \|27\| 6 means \$2710 for company *A* and \$2760 for company *B*

(a) Find the median and the interquartile range of the monthly salaries of employees in company A. **[2**

The lower quartile, median and upper quartile for company B are \$2600, \$2690 and \$2780 respectively.

(b) Draw two box and whisker plots in a single diagram to represent the information for the salaries of employees at companies A and B.

(c) Comment on whether the mean would be a more appropriate measure than the median for comparing the given information for the two companies.

(d) An outlier is defined as any data value which is more than 1.5 times the interquartile range above the upper quartile or 1.5 time the interquartile range below the lower quartile. How high or low must a salary be in order to be classified as an outlier.

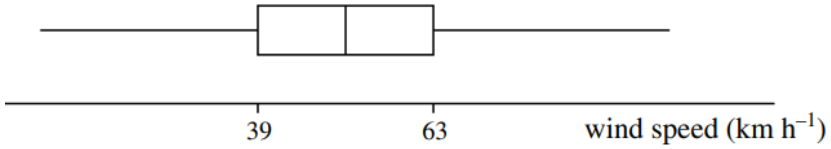
Q14

(a) During January the numbers of people entering a store during the first hour after opening were as follow:

Time after opening, x minutes	Frequency	Cumulative frequency
$0 < x \leqslant 10$	210	210
$10 < x \leqslant 20$	134	344
$20 < x \leqslant 30$	78	422
$30 < x \leqslant 40$	72	a
$40 < x \leqslant 60$	b	540

Find the values of a and b **[2]**

(b)



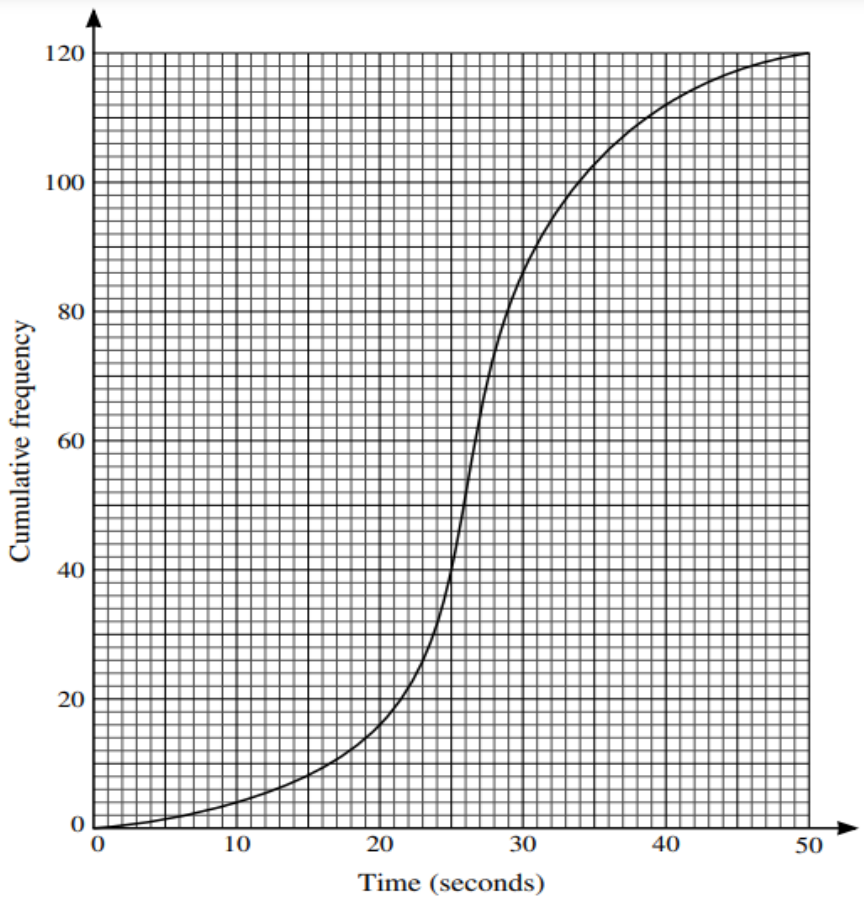
Measurements of wind speed on a certain island were taken over a period of one year. A box and whisker plot of the data obtained is displayed above and the quartiles are as shown. It is suggested that the wind speed can be modelled approximately by a symmetrical distribution. Estimate the mean value. **[1]**

(c) Red Street Garage has 9 used cars for sale. Fair wheel Garage has 15 used cars for sale. It is given that $\sum x = 64$ and $\sum x^2 = 352$, where x is the age of a car in years.

Find the mean and standard deviation of the ages of all 24 cars. **[3]**

Q15

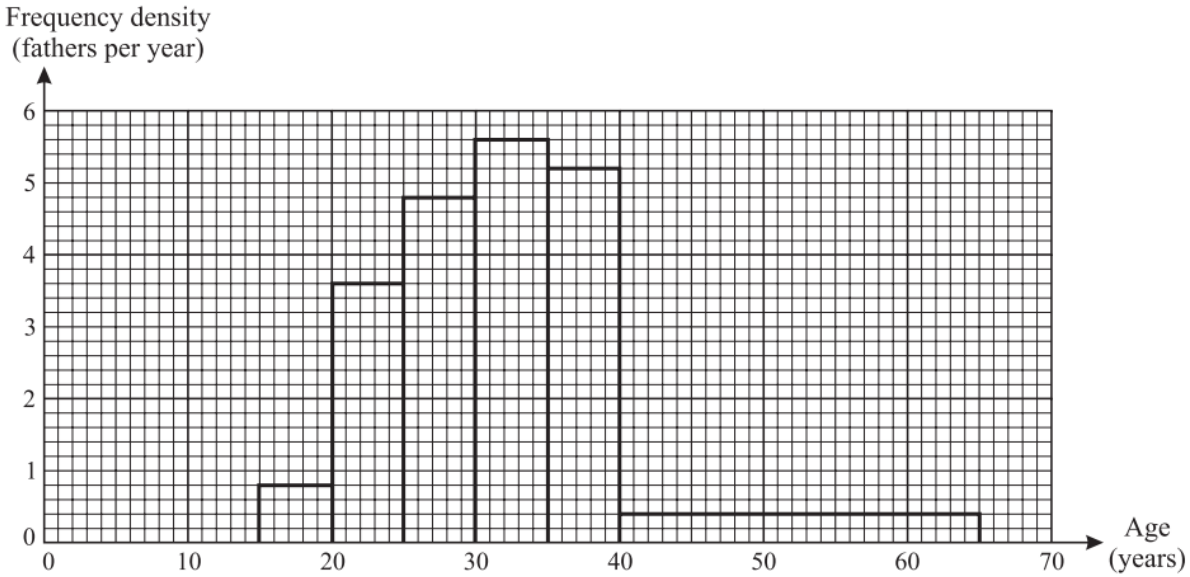
(a) The times taken by children to complete a particular puzzle are represented in the cumulative frequency graph.



(i) Use the graph to calculate the number of children who participated in the puzzle and the maximum time taken.

(ii) 40% of the children took longer than T seconds to complete the puzzle, use your graph to estimate the value of T.

(b) **Each** father in a random sample of fathers was asked how old he was when his first child was born. The following histogram represents the information.



(i) What is the modal age group? **[1]**

(ii) How many fathers were between 25 and 30 years old when their first child was born?

(iii) How many fathers were in the sample?

Q16

The following dataset contains information about the advertising expenditure (in thousand dollars) and the corresponding sales (in thousand units) for a company over 10 months.

Month	1	2	3	4	5	6	7	8	9	10
Advertising Expenditure (x)	2	3	5	7	8	9	10	12	14	16
Sales (y)	4	5	7	10	12	14	15	18	19	22

(i) Calculate the correlation Coefficient (r) and interpret the result

(ii) Find the linear regression equation in the form $y = a + bx$

(iii) Use the regression equation to predict the sales if the advertising expenditure is \$11,000.

(iv) Interpret the slope and intercept of the regression line in the context of this problem.