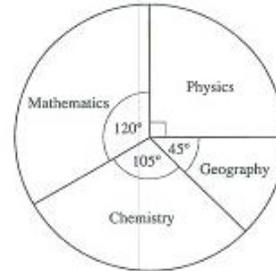


**Topic 18**

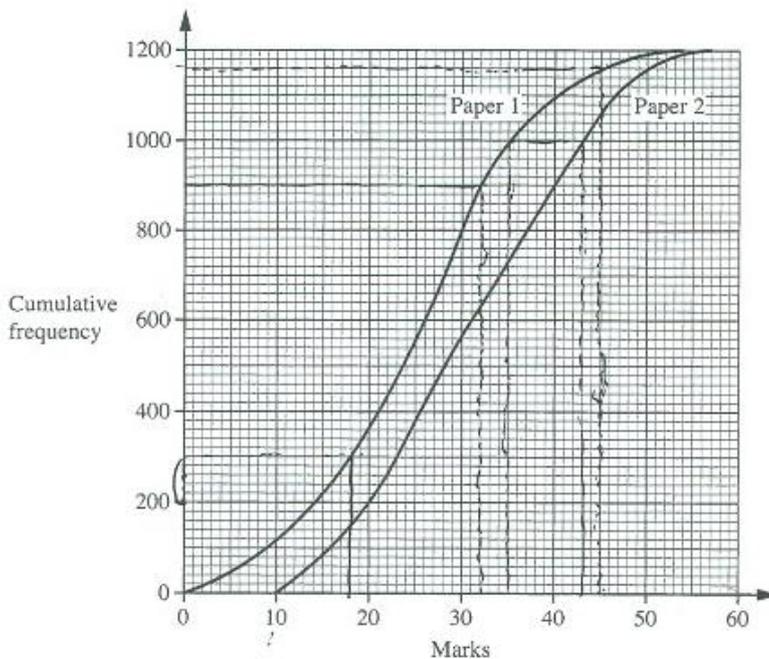
**Data Analysis**

1. The students in a college were asked to vote for their favourite subject. The pie chart represents the number of votes for each subject. Mathematics obtained 25 more votes than Chemistry. Calculate the number of students who took part in the survey.



[2]  
(N2011/P1/Q3)

2. In a Mathematics examination, 1200 students each took two papers. Both papers were marked out of 60. The cumulative frequency curves show the distribution of the marks for the two papers.



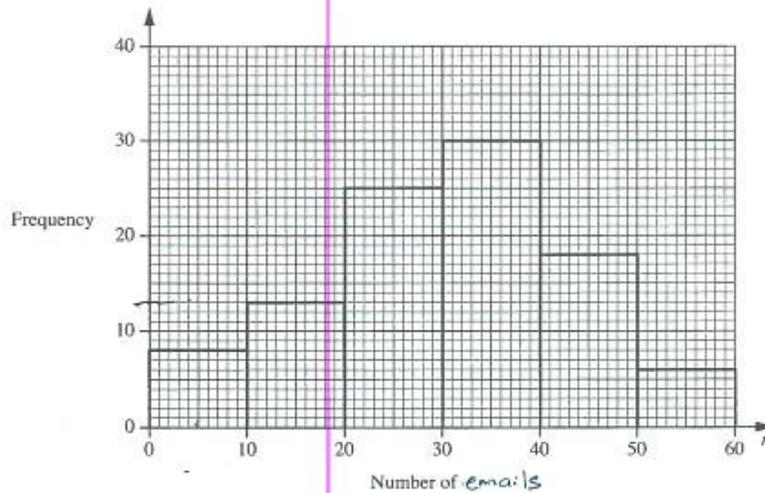
- (a) For Paper 1, estimate the interquartile range. [2]  
(b) Estimate the number of candidates that gained at least 75% of the marks for Paper 2. [2]  
(c) One additional student gained 35 marks in Paper 1, but was absent for Paper 2. Estimate the mark he would have obtained if he had taken Paper 2. [1]  
(d) Which was the more difficult paper? Justify your answer. [1]

(N2011/P1/Q23)

3. A survey was carried out to find the number of emails received in a week by each of a group of students. This information was grouped as follows.

Number of emails ( $n$ )	$0 \leq n < 10$	$10 \leq n < 20$	$20 \leq n < 30$	$30 \leq n < 40$	$40 \leq n < 50$	$50 \leq n < 60$
--------------------------	-----------------	------------------	------------------	------------------	------------------	------------------

The histogram below represents the results of the survey.



- (a) Find the percentage of students who received less than 20 emails in a week. [2]  
 (b) Calculate an estimate of  
 (i) the mean number of emails received in a week, [1]  
 (ii) the standard deviation. [2]

(N2011/P2/Q10a)

4. The stem-and-leaf diagram shows the times, in minutes, taken by some students to complete a task.

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

Key : 3 | 8 represents 38 minutes

For these times, find

- (a) the range, [1]  
 (b) the median. [1]

(N2012/P1/Q4)

TOPIC 18 Data Analysis

5. The table shows the heights of a group of 30 girls.

Height $h$ (cm)	Frequency
$140 \leq h < 150$	4
$150 \leq h < 160$	9
$160 \leq h < 170$	11
$170 \leq h < 180$	6

Calculate an estimate for

- (a) the mean height of the girls,  
(b) the standard deviation of their heights.

[1]

[2]

(N2012/P1/Q12)

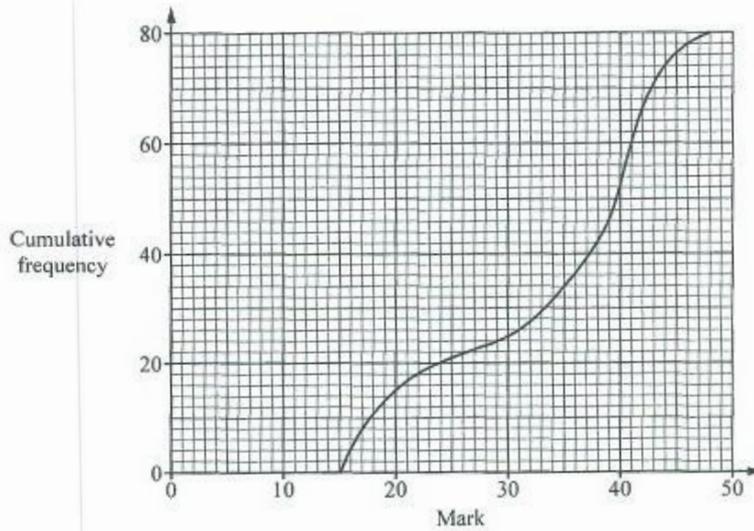
6. The table shows what some pupils chose to do after leaving school.

	Employment	Polytechnic	College	TOTAL
Males	43	28	34	105
Females	29			
TOTAL	72	42		180

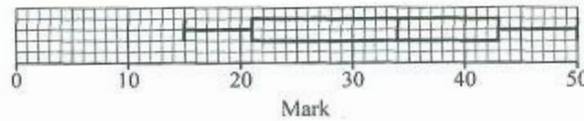
- (a) How many females chose to go to college? [1]  
(b) A pie chart is to be drawn showing the data for the **males**.  
Calculate the angle representing the males choosing polytechnic. [1]  
(c) In which group, males or females, did a greater percentage choose employment, and by how much? [2]

(N2012/P1/Q17)

7. A group of 80 students took a Physics test.  
The cumulative frequency curve below shows the distribution of their marks.



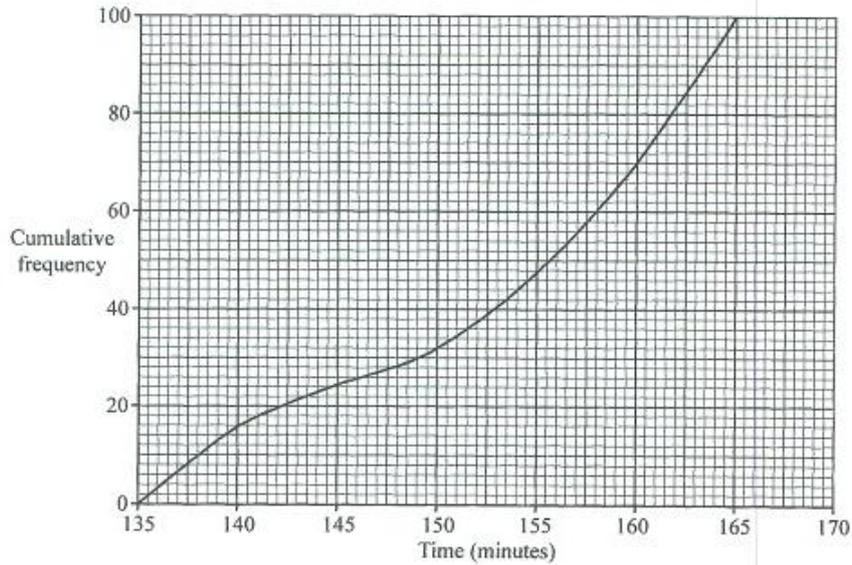
- (a) Use the curve to estimate  
(i) the median mark, [1]  
(ii) the interquartile range. [2]
- (b) The pass mark for the test was 35.  
Estimate the number of students who passed the test. [1]
- (c) The same group of students took a Chemistry test.  
The box-and-whisker plot shows the distribution of their marks.



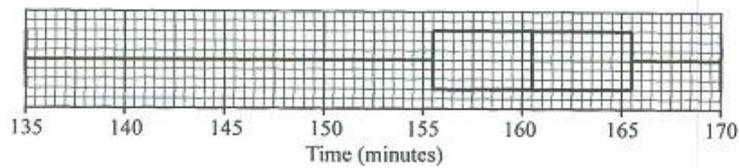
Make two comparisons between the performances of the students in the two tests. [2]  
(N2012/P2/Q10a)

TOPIC 18 Data Analysis

8. The cumulative frequency graph shows the distribution of the times of the first 100 runners to finish a city marathon in 2011.



This box-and-whisker plot represents the distribution of the times of the first 100 runners to finish the city marathon in 2012.



- (a) Use the two diagrams to complete this table for the two marathons.

Year	Lower quartile	Median	Upper quartile	Interquartile range
2011		155.5		
2012	155.5		165.5	10

[3]

- (b) Below are two statements comparing these times over the two years.  
For each one, write whether you agree or disagree, giving a reason for each answer.

Statement	Agree/disagree	Reason
The runners in 2012 were quicker on average.		
The times of the first 25 runners were closer together in 2011.		

[2]  
(N2013/P1/Q22)

9. The heights of a group of 14 students were measured.  
The results are shown in the stem-and-leaf diagram.

15	6				
16	0	2	4	7	8
17	1	2	2	4	8
18	1	3			
19	0				

Key 16 | 2 means 162 cm

- (a) Find the percentage of students who are less than 170 cm tall. [1]  
(b) Find the median height. [1]  
(c) Find the standard deviation of the heights. [2]  
(d) It was discovered that the heights had been measured incorrectly.  
The correct heights were all 5 cm less than those recorded.  
Explain how the median and standard deviation of the heights have been affected by this error. [2]

(N2013/P2/Q10b)

TOPIC 18 Data Analysis

10. Alyssa wants to find out how much time students spend on the Internet. She uses this question on a questionnaire.

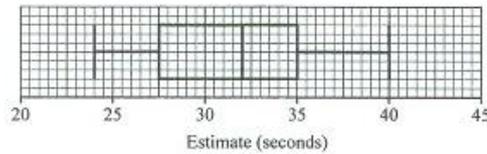
How many hours do you spend on the Internet?  
Tick one box.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1-2	3-4	5-7	over 7

List **two** things wrong with this question.

[2]  
(N2014/P1/Q5)

11. A group of students was asked to estimate a time interval of 30 seconds. The results are represented in the box-and-whisker plot.



- (a) Find the median of the estimates.  
(b) Find the interquartile range of the estimates.

[1]  
[1]  
(N2014/P1/Q10)

12. The table below summarises the times taken by 100 males to complete a 10 km race.

Time ( $t$ minutes)	$30 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$	$60 \leq t < 70$	$70 \leq t < 80$
Frequency	15	32	30	16	7

- (a) What percentage of the males ran faster than 10 km/h? [1]  
(b) Calculate an estimate of [1]  
(i) the mean time, [1]  
(ii) the standard deviation. [2]  
(c) The mean time for females to complete the race was 58.3 minutes and the standard deviation was 9.6 minutes. [2]  
Make two comparisons between the times for males and the times for females.

(N2014/P2/Q10a)

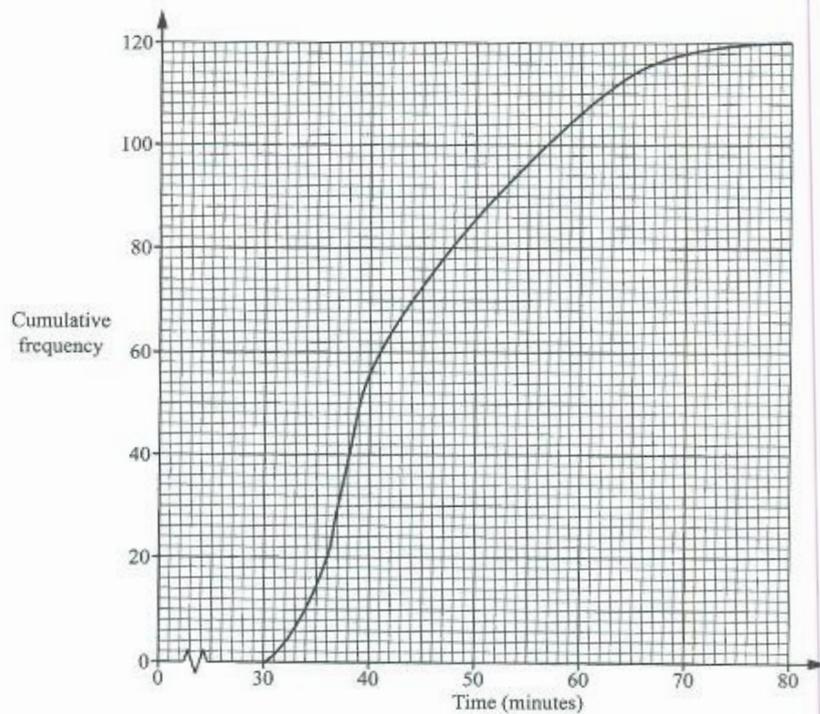
13. Alec has written down five numbers.  
The mean of these numbers is 7, the median is 5 and the mode is 4.  
The largest number is three times the smallest number.

Find the five numbers.

[2]  
(N2015/P1/Q3)

14. Simon weighed six apples.  
The mean mass of the apples was 137 grams.  
The standard deviation of the masses of the apples was 5.48 grams.  
The scales used by Simon were found to be inaccurate.  
The correct mass of each apple was 25 grams more than Simon recorded.  
Write down the correct values for the mean and standard deviation (SD). [2]  
(N2015/P1/Q9)

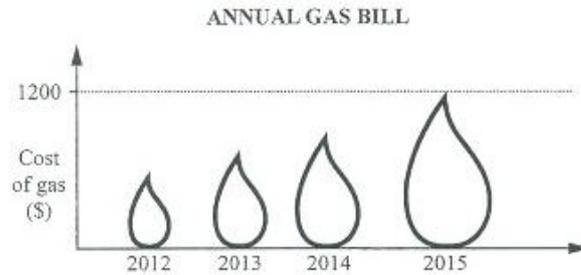
15. The times taken by 120 men to complete a cycle race are recorded.  
The cumulative frequency curve below shows the distribution of their times.



- (a) Use the curve to estimate  
(i) the median time, [1]  
(ii) the interquartile range of the times, [2]  
(iii) the percentage of the men who took at least one hour to complete the race. [2]  
(b) The times taken by 120 women to complete the cycle race had the same interquartile range as the men's times but a higher median.  
Describe how the cumulative frequency curve for the women may differ from the curve for the men. [1]  
(N2015/P2/Q10a)

TOPIC 18 Data Analysis

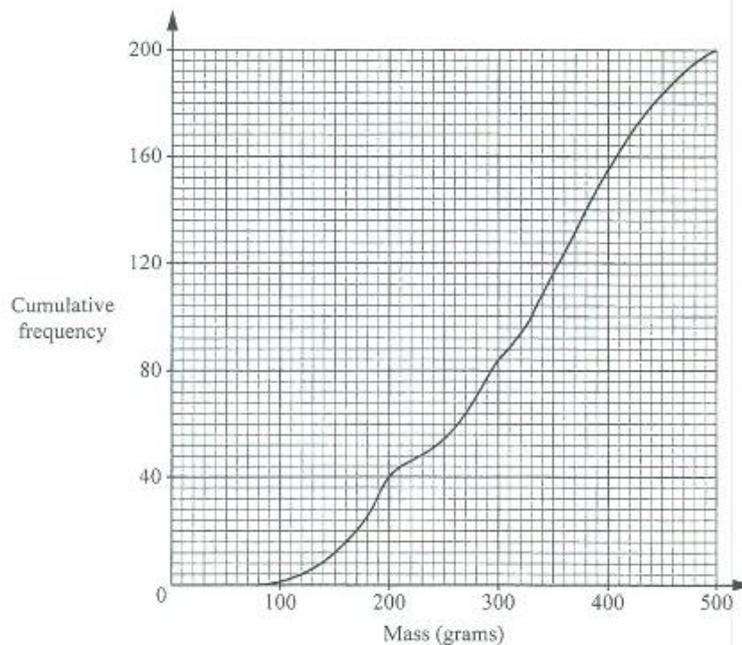
16. Anil draws this graph to show his annual gas bill for each of the last four years.



State one aspect of the graph that may be misleading and explain how this may lead to a misinterpretation of the graph.

[2]  
(N2016/P1/Q8)

17. Some students were asked to estimate the mass, in grams, of sweets in a jar. The cumulative frequency diagram shows the results.



The actual mass of the sweets is 300 grams.

- (a) Find the probability that a student, chosen at random, overestimated the mass. [2]  
(b) Find the number of students who gave estimates within 10% of the actual mass. [2]

(N2016/P1/Q21)

18. A group of 20 students were asked how many minutes they had taken to complete an assignment.  
The results are shown in the stem-and-leaf diagram.



Key: 3 | 2 means 32 minutes

- (a) The teacher expected the assignment to take 30 minutes.  
Find the percentage of students who took longer than expected. [1]
- (b) Find the median time. [1]
- (c) Explain why the mean may not be an appropriate average to use to summarise the times taken by these students. [1]
- (d) Find the standard deviation of the times. [1]
- (e) The standard deviation of the times taken by a second group of students to complete the same assignment was 8.64 minutes.  
Use this information to comment on one difference between the two distributions. [1]
- (N2016/P2/Q9a)
19. The stem-and-leaf diagram shows the masses, in grams, of some apples.



Key: 19 | 4 represents 194 grams

For these masses, find

- (a) the range, [1]
- (b) the median. [1]

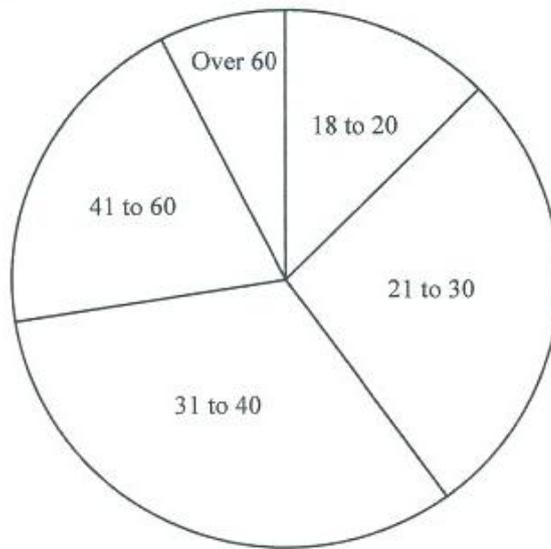
(N2017/P1/Q3)

TOPIC 18 Data Analysis

20. A group of 200 adults took part in a challenge.  
The table below shows the distribution of the times taken to complete the challenge.

Time ( $t$ minutes)	$20 < t \leq 30$	$30 < t \leq 40$	$40 < t \leq 50$	$50 < t \leq 60$	$60 < t \leq 70$
Number of adults	28	84	62	21	5

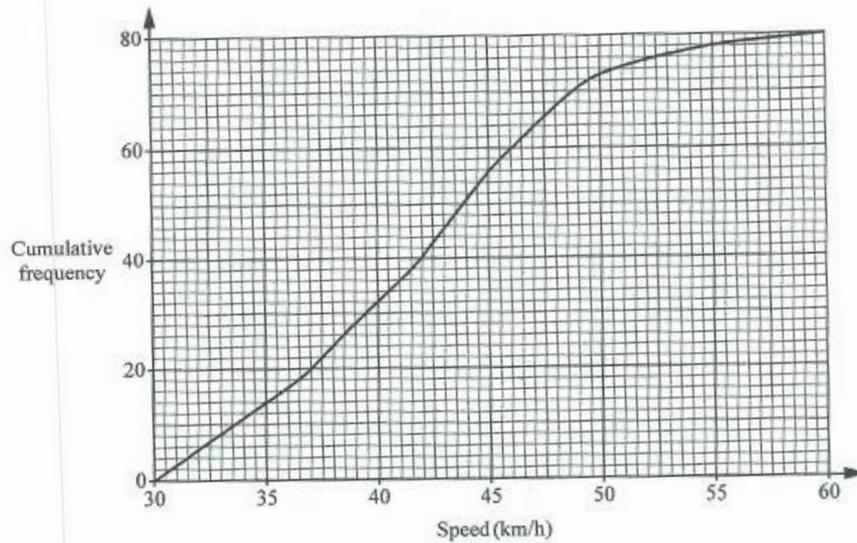
- (a) Calculate an estimate of the mean time. [1]  
(b) Calculate an estimate of the standard deviation of these times. [1]  
(c) This **accurate** pie chart shows the age groups of a second group of adults taking part in the challenge.



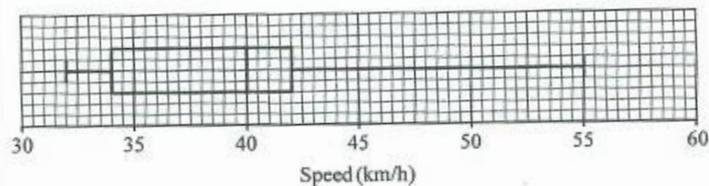
44 of the adults were aged 21 to 30 years old.  
Find the number of adults that were aged 41 to 60 years old.

[2]  
(N2017/P1/Q16)

21. The speeds of 80 cars passing a checkpoint one morning were recorded. The cumulative frequency curve below shows the distribution of the speeds.



- (a) Use the curve to estimate
- (i) the median speed, [1]
  - (ii) the interquartile range of the speeds. [2]
- (b) The speed limit on the road is 50 km/h. Estimate the percentage of the cars that exceeded the speed limit. [2]
- (c) The speeds of 80 cars passing the same checkpoint in the afternoon were also recorded. The box-and-whisker plot shows the distribution of the speeds.

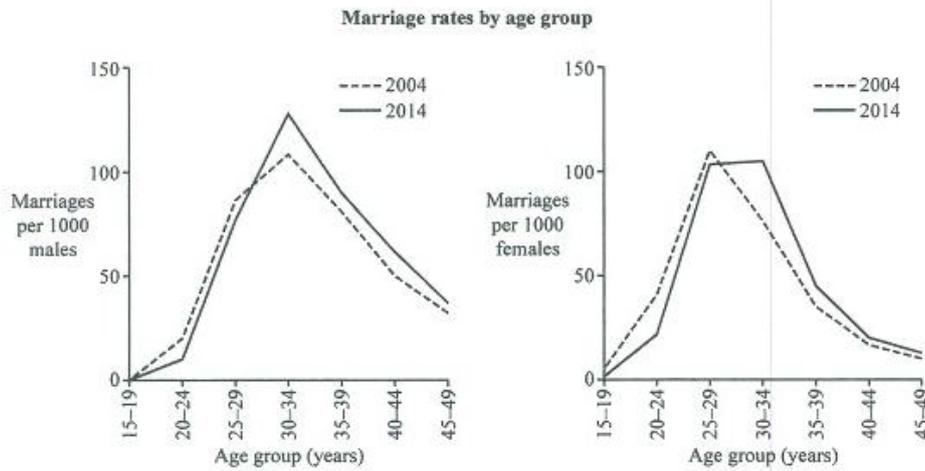


Make two comments comparing the speeds of the cars in the morning and in the afternoon. [2]

(N2017/P2/Q9a)

TOPIC 18 Data Analysis

22.



These charts show the marriage rates for males and females in Singapore for 2004 and 2014.

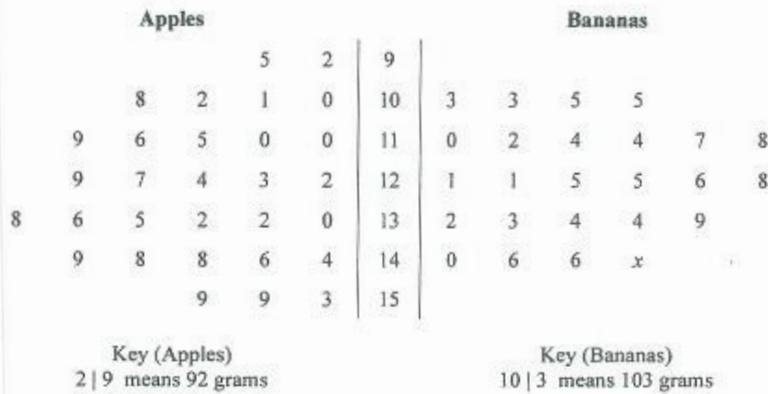
- (a) Make one comment about the ages at which males married in 2014 compared with 2004. [1]
- (b) Huma claims that, in 2014, in general males married at a higher age than females. Do the charts support her claim? Justify your answer with reference to the charts. [1]

(N2018/P1/Q6)

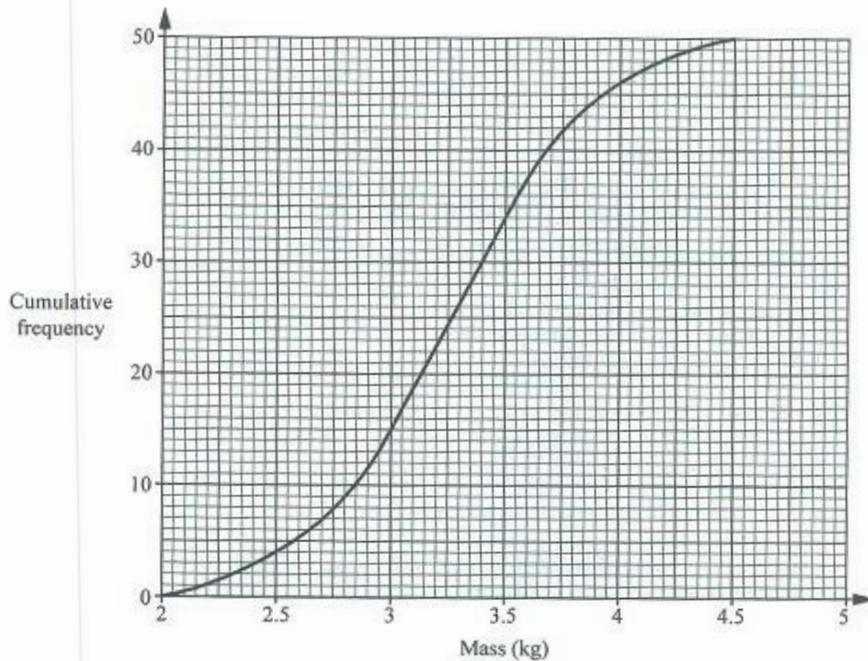
23. The mean height of a group of students is 171 cm. In the group there are 4 more females than males. The mean height of the females is 162 cm. The mean height of the males is 183 cm. Calculate the total number of students in the group.

[4]  
(N2018/P1/Q15)

24. The masses, in grams, of 30 apples and 25 bananas were recorded. The results are shown in the stem-and-leaf diagram.



- (a) Write down the median mass of the apples. [1]
- (b) The range of the masses for the bananas is 44 grams.  
Find  $x$ . [1]
- (c) Make two comments comparing the masses of the apples and the bananas. [2]
- (N2018/P1/Q19)
25. Jiao is researching maternity care in a hospital. She searches the internet to find some information.
- (a) She finds information about the masses of 50 newborn babies born in the hospital. The cumulative frequency curve shows the distribution of the masses.



TOPIC 18 Data Analysis

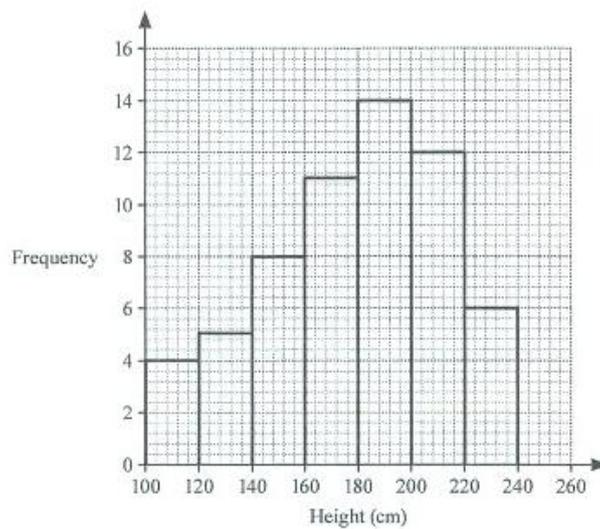
(i) Copy and complete the grouped frequency table for the masses of the babies.

Mass ( $m$ kg)	$2.0 \leq m < 2.5$	$2.5 \leq m < 3.0$	$3.0 \leq m < 3.5$	$3.5 \leq m < 4.0$	$4.0 \leq m < 4.5$
Frequency	4	11			

- (ii) Calculate an estimate of the mean mass of each baby. [1]
- (iii) Calculate an estimate of the standard deviation. [1]
- (iv) Explain why the mean and standard deviation are estimates. [1]
- (v) A newspaper article states that 90% of newborn babies have a mass greater than 2.8 kg. Comment on whether the data from the hospital supports this claim. [2]

(N2018/P2/Q9a)

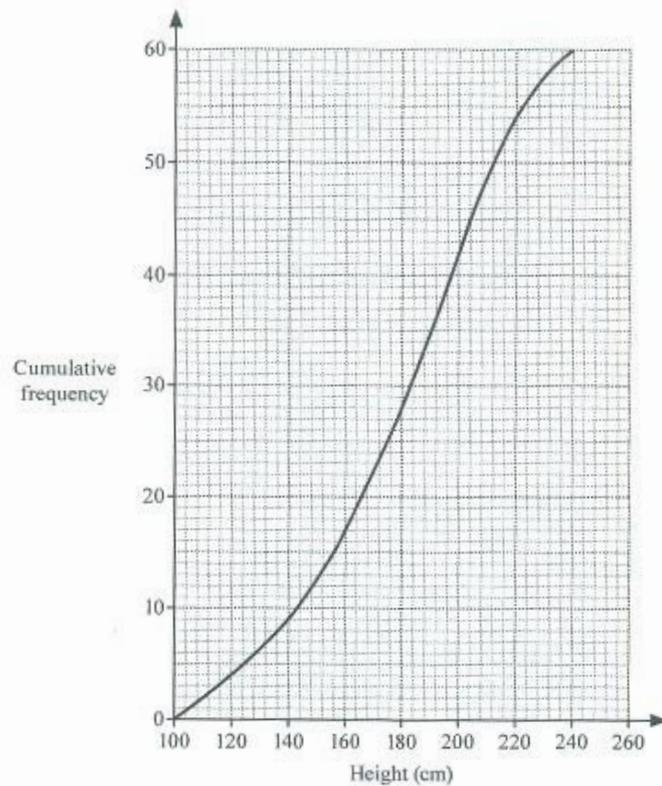
26. (a)



The histogram shows the distribution of the heights of 60 young trees.

- (i) Which interval contains the median height? [1]
- (ii) Calculate an estimate of the mean height of the trees. [1]

(b)



The cumulative frequency graph also represents the heights of the 60 trees. Use the graph to find

- (i) the number of trees with a height of at least 1.92 m, [1]
  - (ii) the interquartile range of the heights. [1]
- (c) Another set of trees has heights with an interquartile range of 36 cm. Explain what this tells us about the second set of trees compared with the first. [1]

(N2019/P1/Q24)

TOPIC 18 Data Analysis

27. Two groups of students entered a science competition .  
Their scores are shown in the stem-and-leaf diagram.



Key: 3 | 5 | 0 means a score of 53 in group A and a score of 50 in group B

- (a) Find the median score for group A. [1]
- (b) Find the range of scores for group B. [1]
- (c) Calculate the percentage of students in group A who scored more than 70. [1]
- (d) Students who scored more than 70 were awarded a merit.  
Students who scored more than 85 were awarded a distinction.  
Use this information to make two comments comparing the performance of the students in the two groups . [2]

(N2019/P2/Q9a)