

Core: Data analysis
Analysis questions

Question 1

Table 1 below shows the home-loan interest rates from 16 lenders for the Standard Rate and the Basic Rate for 12 of the lenders. (Source: InfoChoice 20/02/2002)

Table 1: Home-loan Interest Rates

Lender	Standard Rate (%)	Basic Rate (%)
AMP Banking	6.05	5.49
ANZ Bank	6.07	5.47
Aussie Home Loans	5.95	5.49
AXA Home Loans	5.74	—
Bank of Melbourne	6.07	5.49
Bendigo Bank	6.10	5.59
Commonwealth Bank	6.07	5.56
Credit Union Australia	5.99	—
Education Credit Union	6.05	—
EAI Home Loans	5.90	5.55
IJ Hooker Home loans	6.07	5.62
Loan Corporation	5.65	5.45
Melbourne Credit Union	5.35	—
National Australia Bank	6.06	5.56
St George Bank	6.07	5.49
Wizard Mortgage	5.72	4.89

Part 1

- a. The mean and standard deviation for Standard home-loan rates are given below (correct to 2 decimal places):

mean = 5.93%

standard deviation = 0.21%

It follows that 95% of the Standard home-loan rates would be expected to be between _____% and _____%

[1 mark]

- b. Using Table 1 above, determine the mean and standard deviation for the Basic rate (correct to 2 decimal places).

[2 marks]

- c. Write a sentence comparing the centres of the two interest rate distributions.

[1 mark]

The standardized (z score) rate for the Commonwealth Bank's Standard Rate is 0.67

- d. Calculate the z score for the Commonwealth Bank's Basic Rate. [1 mark]
- e. In comparison with the other lenders, is the Commonwealth Bank more competitive with its Standard or Basic Rate? Justify your answer. [2 marks]

Part 2

In an attempt to predict the Basic rate for the four lenders not listed, a least squares regression line is used. Its equation is:

$$\text{Basic rate} = 0.79 \times (\text{Standard rate}) + 0.72$$

- a. In this investigation, the independent variable is the _____ rate. [1 mark]
- b. Using the regression line above, what would be the Basic rate for the Education Credit Union? [1 mark]
- c. Using the regression line above, if the Basic rate for a new Credit Union (GRAND-IN-A-HAND) was 4.8%, what would its Standard rate be? [2 marks]

Part 3

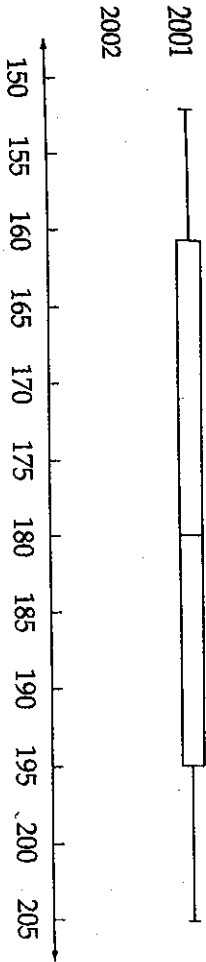
Table 2 shows the prices of 11 new houses advertised by a home building company.

Table 2: New Houses Price List for NEW HOME INC.

House Size (m ²)	2001 Price (\$)	2002 Price (\$)
28.1	161 000	159 000
31.2	172 000	166 000
31.4	152 000	159 000
32.7	161 000	161 000
34.7	185 000	183 000
35.1	180 000	166 000
36.2	177 000	166 000
37.8	190 000	188 000
39.7	195 000	193 000
40.9	195 000	182 000
42.2	205 000	198 000

- a. The box plot below shows the distribution of 2001 new house prices. Use the data in Table 2 to draw a properly scaled box plot displaying the distribution of 2002 new house prices. [3 marks]

Prices in thousands of dollars



- b. Describe the shapes of the 2 box plots. [2 marks]
 c. Compare the medians of the two distributions. [1 mark]
 d. Compare the Inter Quartile Ranges of the two distributions. [1 mark]
- Total = 15 marks

Question 2

The data in Table 1 is based on a survey of students, teachers and parents on the issue of mobile phones. Each person was asked the following question:

“Should students be permitted to bring mobile phones to school?”

Table 1: Mobile Phone Usage in Schools

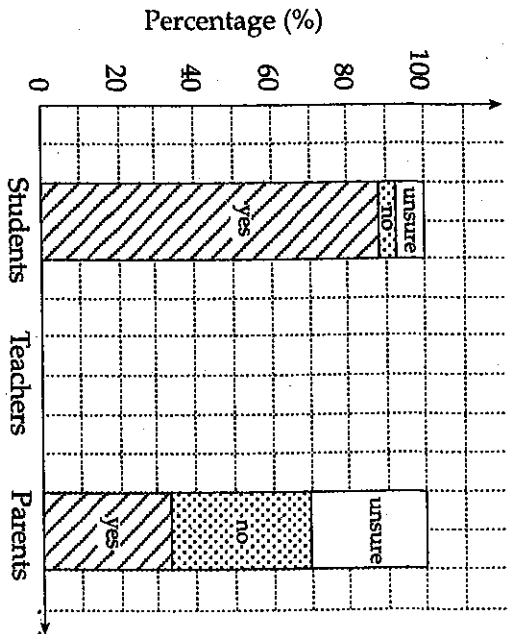
	Yes	No	Unsure
Students	860	40	80
Teachers	22	32	14
Parents	30	32	28

Part 1

- a. Complete the following sentences: [2 marks]
- The number of teachers who responded ‘No’ to bringing mobile phones to school was: _____ [1 mark]
 - The number of students surveyed was: _____ [2 marks]
- b. Of those people who said ‘Yes’, what percentage were teachers? (Give your answer correct to 2 decimal places) [1 mark]

Part 2

The following percentage segmented bar chart shows the response to the question: “Should students be permitted to bring mobile phones to school?”



- a. Draw in the percentage segmented bar for the teachers in the graph above. [3 marks]
- b. Comment on the statement: [1 mark]
- “When comparing teachers’ and students’ views on mobile phones, the number of students (40) who said ‘No’ to bringing mobile phones to school was more than number of teachers (32) who said ‘No.’”

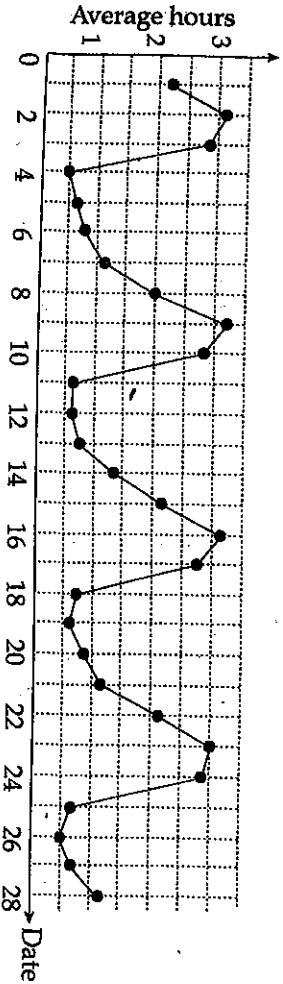
Part 3

In another study of the use of mobile phones, information was collected on the average hours used by students in February, 2002. The data is shown in Table 2.

Table 2 : Average Hours of Students' Use of Mobile Phones (Feb. 2002)

Date	Day	Average Hours Used	Date	Day	Average Hours Used
1/02	Fri	2.2	15/02	Fri	2.1
2/02	Sat	3.1	16/02	Sat	3.3
3/02	Sun	2.9	17/02	Sun	2.7
4/02	Mon	0.5	18/02	Mon	0.7
5/02	Tue	0.6	19/02	Tue	0.6
6/02	Wed	0.8	20/02	Wed	0.9
7/02	Thr	1.2	21/02	Thr	1.1
8/02	Fri	2.0	22/02	Fri	2.1
9/02	Sat	3.2	23/02	Sat	3.0
10/02	Sun	2.8	24/02	Sun	2.9
11/02	Mon	0.6	25/02	Mon	0.7
12/02	Tue	0.6	26/02	Tue	0.5
13/02	Wed	0.7	27/02	Wed	0.7
14/02	Thr	1.3	28/02	Thr	1.2

The time series graph for the average hours of mobile phone use is given below:



- a. Describe the time series graph. [1 mark]
 b. Comment on phone usage in February. [1 mark]

- c. Complete the tables below:
 i. Three-point moving average smoothing

2/02	3/02	4/02	5/02	6/02	7/02	8/02	9/02	10/02	11/02	12/02
2.7	2.2		0.6	0.9	1.3	2.1	2.7	2.2	1.3	

- ii. Five-point median smoothing

3/02	4/02	5/02	6/02	7/02	8/02	9/02	10/02	11/02	12/02	13/02	14/02
2.2	0.8		0.8	1.2	2.0	2.0	2.0	0.7	0.7	0.7	

- d. What is the most appropriate number of points to use to smooth this data? [2 marks]

Total = 14 marks

Question 3

An airline company based out of Melbourne (Melair) is reviewing its prices and considering flying to other Australian cities. Table 1 below gives the destinations of 12 cities that Melair flies to along with the distance in kilometres from Melbourne to each city and the cost of a one-way flight

Table 1: Cost of a one-way ticket from Melbourne

City	Distance from Melbourne (km)	Cost of one-way flight
Adelaide	732	\$87
Brisbane	1697	\$234
Broken Hill	857	\$189
Cairns	3109	\$379
Canberra	651	\$139
Darwin	3773	\$589
Gold Coast	1746	\$249
Hobart	254	\$134
Katherine	3457	\$612
Perth	3453	\$399
Sydney	876	\$77
Townsville	2763	\$339

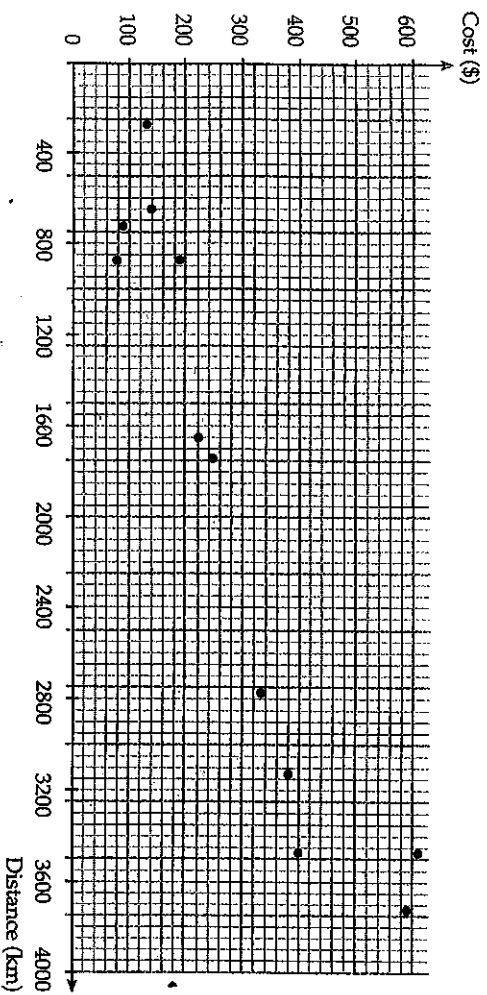
Part 1

- a. What is the range for the distances travelled from Melbourne to the 12 cities given in the table above? [1 mark]

- b. Find the mean cost of a one-way flight for the 12 cities given in the table above (to the nearest dollar). [1 mark]
- c. The company would like to advertise that the mean cost of a one-way ticket to any city that they fly is less than \$270. If they added another city to their destinations, what would the cost of this flight to this city have to be for this claim to be correct? [2 marks]

Part 2

The following scatterplot shows the cost versus distance of one-way travel from Melbourne.

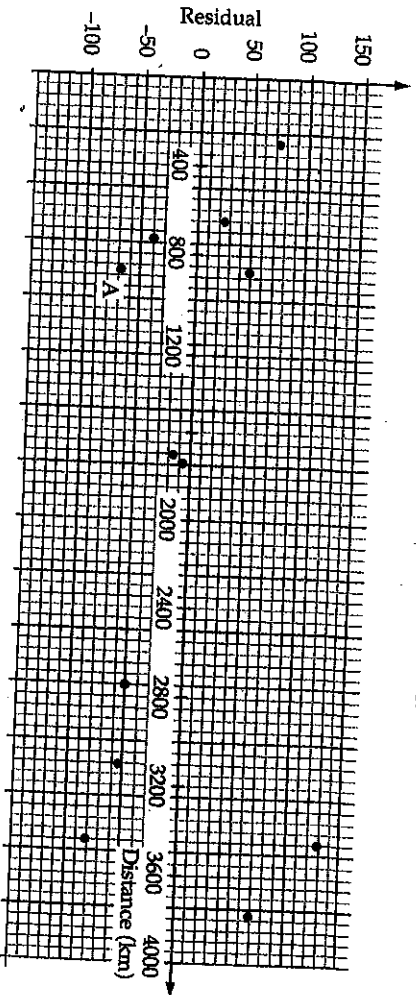


The least squares regression line that would enable the cost to be predicted from the distance is:

$$\text{Cost} = 30 + 0.13 \times \text{distance}$$

- a. Draw this line on the scatterplot above. [2 marks]
- b. What does the 0.13 represent in the consideration of the cost of a one-way flight? [1 mark]
- c. What does the 30 represent in the consideration of the cost of a one-way flight? [1 mark]
- d. The airline decides to include Alice Springs as one of its destinations. Alice Springs is 2270 km from Melbourne. Using the least squares regression line, estimate the cost of a one-way ticket to Alice Springs (to the nearest dollar). [1 mark]
- e. The value of the Pearson's product moment correlation coefficient r is 0.9313 [1 mark]
- i. Comment on what this number shows. [1 mark]

- ii. From this information, Melair can conclude that _____% of the variation in the cost of a one-way ticket can be explained by the variation in the distance travelled. [1 mark]
- f. The following is the residual plot for the data above. [1 mark]



- i. Point A on the residual plot represents the residual for Sydney. What does this residual represent? [1 mark]
- ii. Calculate the residual for Darwin. [2 marks]
- iii. What features of the residual plot support the assumption that the relationship between cost and distance is approximately linear? [1 mark]

Total = 15 marks

Question 4

Part 1

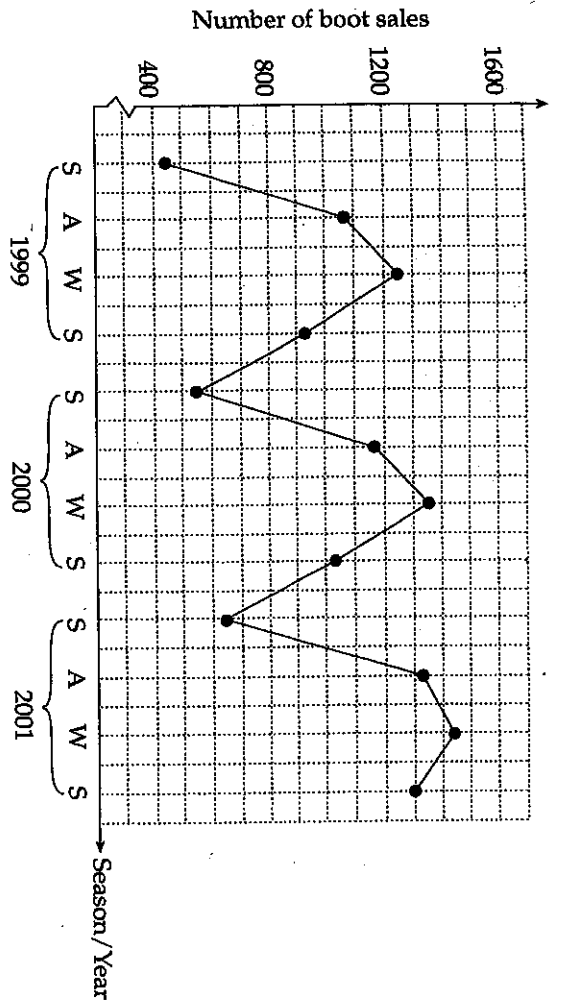
The following data is a record of quarterly sales for "The Boot Company". Table 1 gives the number of boot sales in each of the four seasons for the years 1999 to 2001.

Table 1: Boot sales for 1999–2001 per quarter

Year/Season	Summer	Autumn	Winter	Spring	Average
1999	446	1085	1241	920	923
2000	541	1180	1356	1035	1028
2001	659	1324	1450	1299	
Seasonal index	0.52	1.15		1.03	

- a. What is the average number of sales per season for the year 2001? [1 mark]
- b. What is the seasonal index for Winter? [1 mark]
- c. What does the seasonal index of 0.52 indicate about the sale of boots during the summer? [1 mark]

d. The numbers of boot sales are shown in the time series plot below:



Describe the time series.

[2 marks]

Part 2

a. The Boot Company would like to predict sales for the four seasons in the year 2002. To do this they deseasonalise the data, as shown in Table 2 below.

Table 2: Deseasonalised data

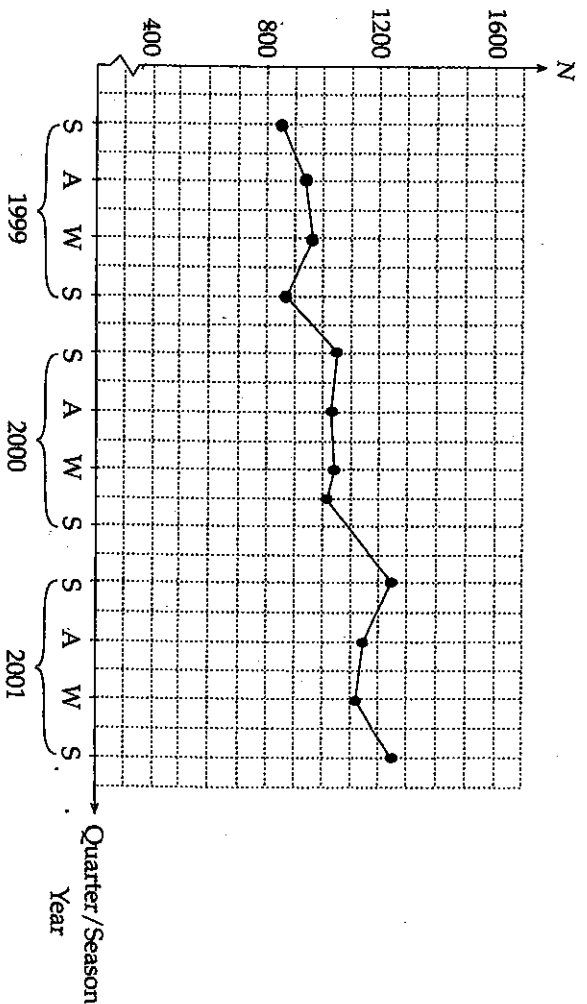
Year/Season	Summer	Autumn	Winter	Spring
1999		943	955	893
2000	1040		1043	1005
2001	1267	1151	1115	1261

Complete the table for the Summer of 1999 and Autumn 2000.

[2 marks]

b. The time series plot below shows the deseasonalised data for boot sales. On this time series plot, draw the three-median line.

[3 marks]



Part 3

a. Another trend line is calculated which is given by:

$$N = 839 + 32 \times \text{time (time in quarters)}$$

Using this trend line, calculate the value of N for Winter 2002.

[2 marks]

b. Using this value, estimate the number of boot sales for Winter 2002. [1 mark]

Part 4

Instead of deseasonalising the data, The Boot Company considers a technique known as smoothing. Moving mean smoothing or median smoothing can do this.

- a. What would be the most appropriate number of points for The Boot Company to use for moving mean smoothing? [1 mark]
- b. What is the advantage of using median smoothing? [1 mark]

Total = 15 marks

Question 6

Table 1 is the tally of medals from countries that received 2 or more gold medals at the 2002 Winter Olympic Games.

Table 1: Medal Tally of 2002 Winter Olympic Games

Country	Gold	Silver	Bronze
Australia	2	0	0
Austria	2	4	10
Canada	6	3	8
China	2	2	4
Croatia	3	1	0
Finland	4	2	1
France	4	5	2
Germany	12	16	7
Italy	4	4	4
Korea	2	2	0
Netherlands	3	5	0
Norway	11	7	6
Russian Federation	6	6	4
Spain	2	0	0
Switzerland	3	2	6
USA	10	13	11

Part 1

- a. Which country won the most bronze medals? [1 mark]
- b. Which country won the most medals overall? [1 mark]

The following data are the scores (out of 50) of two Further Mathematics classes. The back-to-back stem plot is shown below.

Class A

19 25 35 35 30 36 42 20 38 39
 47 28 43 45 26 37 42 43 40 26
 48 40

Class B

5 41 9 31 11 8 42 7 44 9
 47 45 13 19 32 38 37 23 24

Class A

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

Class B

Part 1

- a. Describe the shapes of the two distributions – Class A and Class B. [2 marks]
- b. State the median for each distribution. [2 marks]

Part 2

The scores (out of 20) of a third Further Mathematics Class are shown below.

Class C

2 9 11 5 16 17 12 7 15 10
 12 16 17 18 19 16 15 13 14

- a. Using appropriate split stems, display the data on a stem and leaf plot. [3 marks]
- b. Calculate the Interquartile Range for Class C. [2 marks]
- c. State, with explanation, whether the distribution has any outliers. [2 marks]

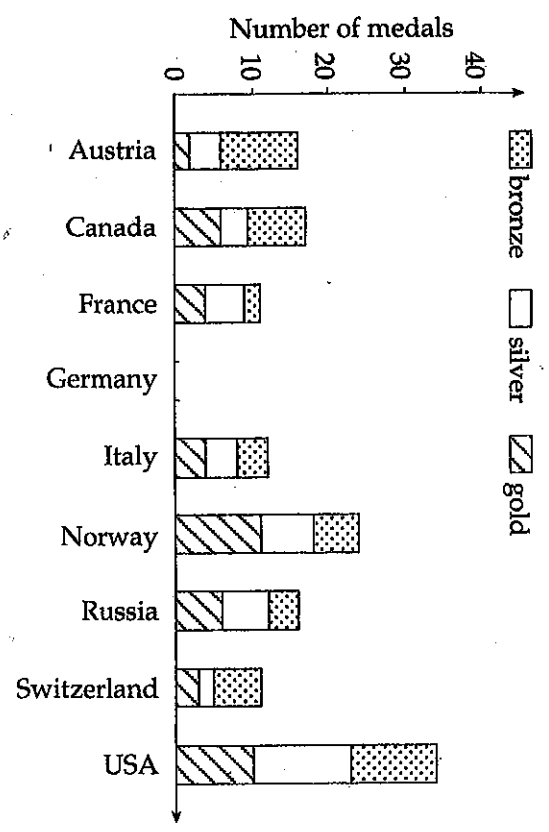
Part 3

One student is to be selected from each of the three Further Mathematics classes; Class A, Class B and Class C.

Describe briefly how you would choose one student from each class using a simple random sample. [2 marks]

Total = 13 marks

The segmented bar chart below displays the data of the countries that received more than 10 medals overall.



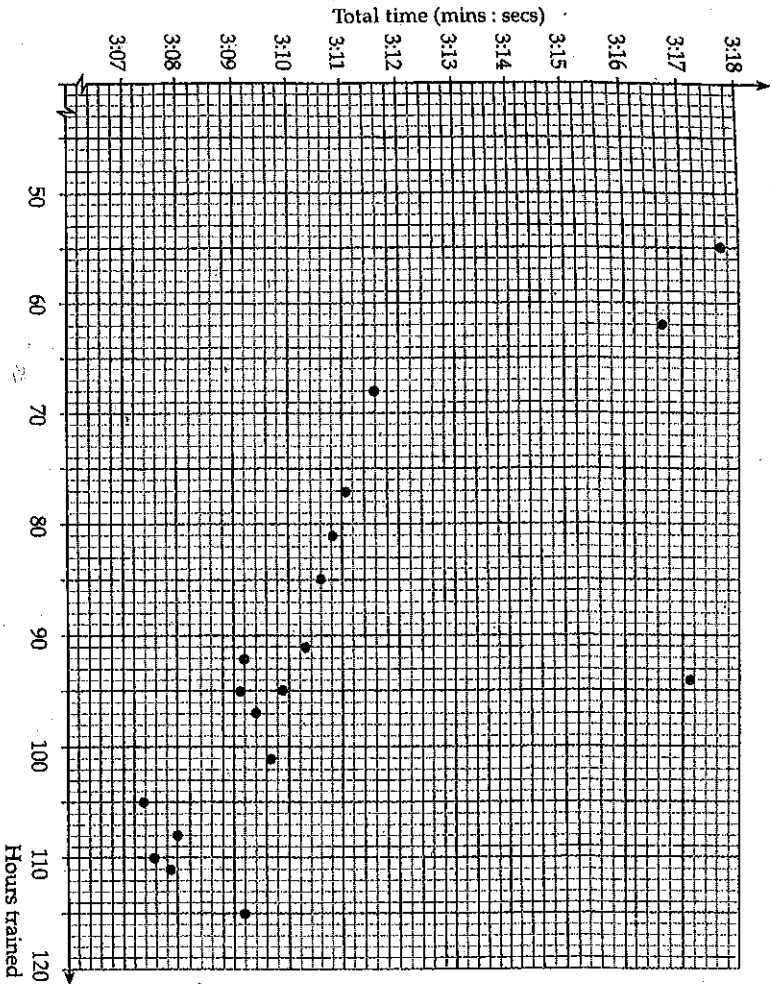
Complete the segmented bar graph for Germany. [2 marks]

The data in Table 2 shows the top team for each country that competed in the 4-Man Bob Sleigh race at the 2002 Olympic Games along with the hours they trained just prior to the Olympics and their total time for the four heats in the Olympic Games.

Table 2: Hours trained and total time (4 heats) for 4-Man Bob Sleigh race

Country	Hours trained (just prior to the Games)	Total time (in minutes and seconds) (4 heats in the Olympics)
Germany	105	3:07.51
USA	110	3:07.81
Switzerland	108	3:07.95
France	111	3:07.95
Latvia	95	3:09.06
Russia	92	3:09.15
Canada	105	3:09.17
Great Britain	97	3:09.37
Austria	101	3:09.57
Czech Republic	95	3:09.93
Netherlands	91	3:10.38
Poland	85	3:10.73
Italy	81	3:10.96
Japan	77	3:11.17
Romania	68	3:11.66
Brazil	62	3:16.73
Monaco	94	3:17.19
Taipei	55	3:17.76

The scatterplot below shows the Total time versus Hours trained for the data in Table 2.



a. One point is an outlier. Which country does it represent? [1 mark]

b. The least squares regression line that fits this data is in the form:

$$y = a + bx$$

where x is the number of hours trained and the total time is 3 minutes y seconds.

Find the values of a and b . [2 marks]

Draw this line on the scatterplot. [2 marks]

Comment on the appropriateness of this linear model for the data. [1 mark]

The Brazilian team believes that if they train for 200 hours just prior to the 2006 Winter Olympic Games they have a good chance of winning the gold medal.

Comment on this statement. [1 mark]

Part 4

The winner of the gold medal in the Women's Aerial Free-Style Skiing was Alisa Camplin from Australia. The scores for a Women's Aerial Free Style Competition are given below:

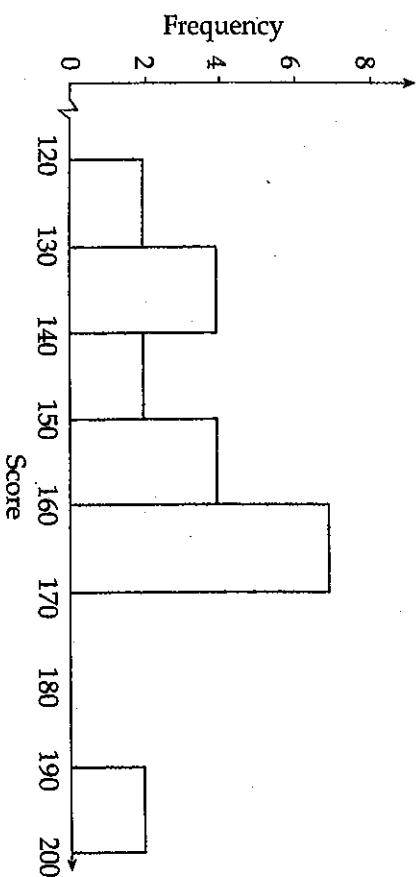
193.47	190.02	189.26	188.37	185.23	174.24
170.54	169.38	164.92	153.88	147.31	135.96
167.96	152.07	151.23	148.82	137.80	136.77
130.31	125.47	125.20	171.81	162.56	151.09
172.11	178.62	163.91	175.04	166.34	162.62

a. Complete the frequency table below for the data. [2 marks]

Class interval	Frequency
120-	2
130-	4
140-	2
150-	4
160-	7
170-	
180-	
190-	2
Total	

b. The data is displayed in the histogram below:

[2 marks]



i. Complete the histogram for 170 to 190. [2 marks]

ii. Describe the shape of the histogram. [1 mark]

[2 marks]

[1 mark]

Total = 16 marks