

# Unit 2 Test: Data Analysis

## Interpreting Sets of Data and The Normal Distribution and Correlation

Remember: these are HSC-type questions.

**Time allowed: 1 hour 10 minutes Total marks: 50**

### Part A (Suggested time: 30 minutes)

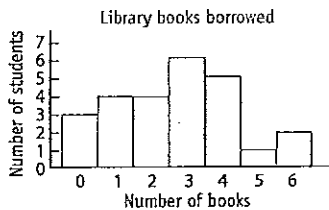
Choose the correct answer (A, B, C or D) for each question. One mark each

**1** The mean of a set of eight scores is 15. A new score of 6 is added to the set. What is the new mean?  
 A 15.75                      B 14.75  
 C 14.25                      D 14

**2** In a normal distribution, the mean is 66 and the standard deviation is 11. Approximately what percentage of scores lie between 55 and 88?  
 A 68%                      B 74.5%  
 C 81.5%                      D 95%

**3** The correlation coefficient between two sets of data is 0.2. Describe the correlation.  
 A weak                      B moderate  
 C moderately strong      D very strong

**4** A survey was taken of the number of books students had borrowed from a school library. The results are displayed in the histogram.



What is the mode?  
 A 3                      B 4  
 C 5                      D 6

**5** Referring to the histogram in question 4, how many books have been borrowed?  
 A 21                      B 25  
 C 67                      D 70

**6** In an exam, Maree scored 87%. The mean was 63.2% and the standard deviation was 8.75%. What is Maree's z-score?  
 A 2.12                      B 2.32  
 C 2.52                      D 2.72

**7** A survey of 21 families was taken and the number of children in each family was recorded. The results were: 6 2 1 4 3 2 3 0 2 2 4  
 3 1 3 5 21 2 3 4 2 3  
 What is the outlier?  
 A 6                      B 0  
 C 21                      D 3

**8** What is the median of the scores:  
 9 3 4 4 8 11 6 2?  
 A 4                      B 5  
 C 6                      D 8

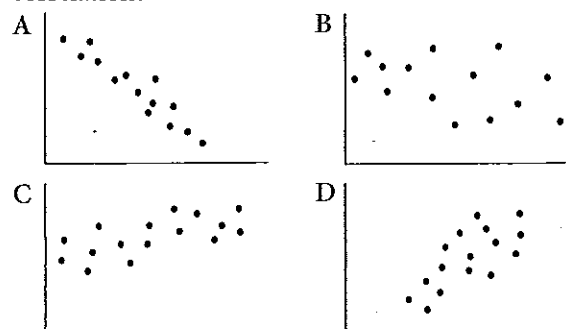
**9** Joshua achieved a z-score of 2.4 in his music exam. His raw mark was 72% and the standard deviation was 6%. What was the mean?  
 A 57.6%                      B 63.6%  
 C 80.4%                      D 86.4%

**10** The results by two classes in the same exam are shown in the back-to-back stem-and-leaf plot. By how much does the median of the class Y exceed the median of class X?

Exam results	
X	Y
9 8 5 0	5   1 4 6
8 6 4 3	6   0 3 7 7 8
8 6 2 2 2	7   1 4 5 5 6 7 9 9
7 6 4 3 1 0	8   3 5 7 8
5 0	9   1

A 3                      B 5  
 C 1                      D 1.5

**11** Which scatterplot shows a strong negative correlation?



**12** Ken, Barbara and Jess are all in different years at the same school. In the end of year exams, Ken scored 75% in maths. The mean mark was 81% and the standard deviation 7.5%. Barbara scored 62% in her maths exam. The mean was 67% and the standard deviation 6.25%. Jess scored 68%, the mean mark was 74% and the standard deviation was 12%. Who achieved the best result?

- A Ken
- B Barbara
- C Jess
- D both Ken and Barbara

**13** Clients at a travel office were surveyed as to whether they wished to holiday in Australia or overseas. The results are given in the two-way table.

	Men	Women
Australia	28	56
Overseas	42	64

What percentage of men wanted to holiday in Australia?

- A  $66\frac{2}{3}\%$
- B 50%
- C 40%
- D  $33\frac{1}{3}\%$

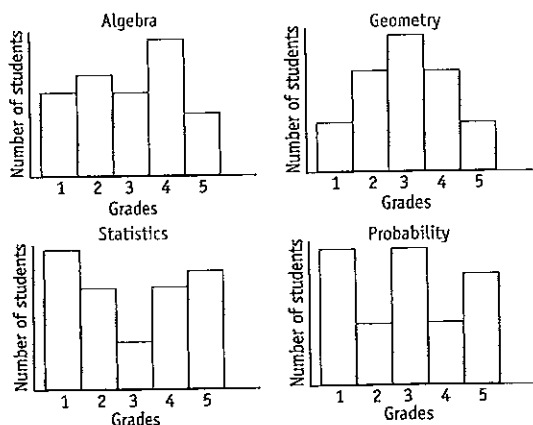
**14** Brooke's z-score for her English exam was 3. If 48 000 students sat for the exam and the results were normally distributed, approximately how many students did better than Brooke?

- A 72
- B 120
- C 1200
- D 7200

**15** The scores 143 157 191 156 184 176 138 145 152 166 171 162 have a mean of 161.75 and a standard deviation of 15.82. How many of the scores are within one standard deviation of the mean?

- A 4
- B 5
- C 7
- D 8

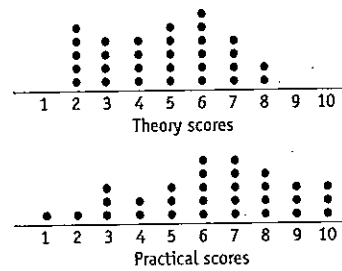
**16** A mathematics test consisted of four different topic areas. The results for each topic were given as a grade from 1 to 5, with 5 being the highest, and the results are displayed below.



Which topic area had grades with the lowest standard deviation?

- A algebra
- B geometry
- C statistics
- D probability

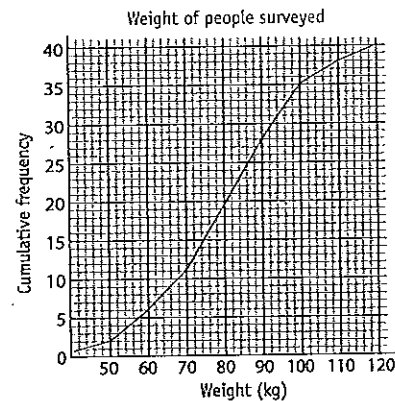
**17** A theory and a practical test were both given to the same set of students in the same subject. The results are shown on the dot plots, using the same scale.



Which statement is correct?

- A The mean and standard deviation were both higher in the theory test.
- B The mean and standard deviation were both higher in the practical test.
- C The mean was higher in theory but the standard deviation was higher in the practical test.
- D The mean was higher in the practical but the standard deviation was higher in the theory test.

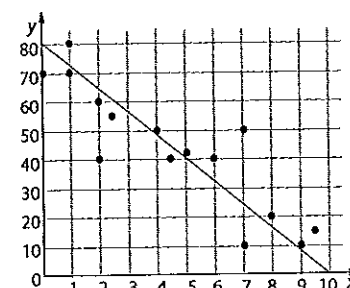
**18** A cumulative frequency polygon has been drawn as the result of a survey of people's weight.



What is the interquartile range?

- A 15 kg
- B 25 kg
- C 35 kg
- D 40 kg

**19** A line of fit has been drawn on the scatterplot as shown.



What is the equation of the line of fit?

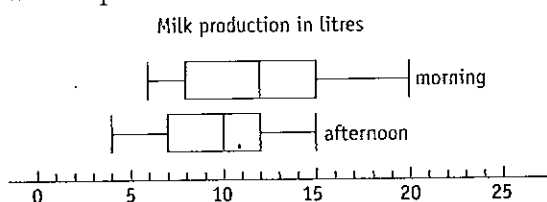
- A  $y = 8x + 80$
- B  $y = 8x - 80$
- C  $y = -8x + 80$
- D  $y = -8x - 80$

- 20** In a normal distribution, the mean is 16.5 and the standard deviation 4.5. Between what values will approximately 95% of the scores lie?
- A 16.5 and 30                      B 12 and 21  
C 7.5 and 25.5                      D 21 and 30

**Part B** (Suggested time: 40 minutes)

Show all working. 30 marks

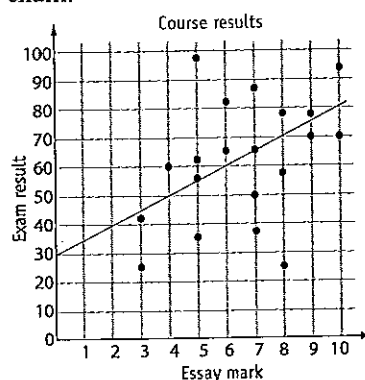
- 21** A herd of dairy cows was tested to see what their milk production in litres was in both the morning and afternoon milkings on a particular day. The box-and-whisker plots were drawn to illustrate the results.



- a What is the median milk production in the afternoon? 1 mark
- b Approximately what percentage of cows produced more than 8 litres in the morning? 1 mark
- c What is the interquartile range for the afternoon milking? 1 mark
- d Describe any similarities and differences that can be seen from the displays. 2 marks

- 22** In a normal distribution the mean is 113 and the standard deviation 12.
- a Draw up a table showing the scores equivalent to z-scores of -3, -2, -1, 0, 1, 2 and 3. 1 mark
- b What is the median? Justify your answer. 2 marks
- c Approximately what percentage of scores lie between 89 and 137? 1 mark
- d Approximately what percentage of scores lie between 101 and 149? 2 marks
- e The scores are marks in an exam out of 150. 12 000 students sat for the exam. Approximately how many students would have scored full marks? 2 marks

- 23** A teacher has prepared a scatterplot, and drawn a line of fit, of the scores by students in an essay and in an exam.



- a What did the student who received an essay mark of 4, score in the exam? 1 mark
- b Briefly describe the correlation between essay marks and exam marks. 1 mark
- c Nathan said: 'Everyone who did well in the essay also did well in the exam.' Give an example to show that Nathan was wrong. 1 mark
- d Find the equation of the line of fit. 2 marks
- e Natalie received 7 for her essay but was away for the exam. The teacher used the line of fit to give her an estimate. What was Natalie's estimated mark? 1 mark

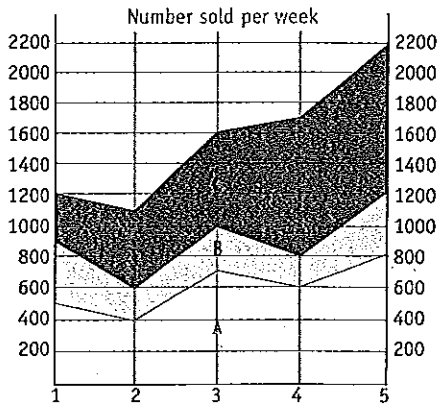
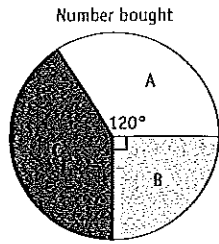
- 24** Nell and the other students in her class did a test. The marks were:

78 56 64 58 69 73 79 59 63  
82 57 77 63 65 56 70 62 66

- a Find the mean. 1 mark
- b Find the standard deviation ( $\sigma_n$ ) to one decimal place. 1 mark
- c Nell scored 73. What is her z-score (to one decimal place)? 1 mark
- d In another test, Nell scored 68. The mean was 60 and the standard deviation 12.5. Which was Nell's better result? 2 marks

25

Six weeks before Christmas a trader bought three different brands of gift packs and sold them at a market over five weekends. The sector graph shows how the 9000 gift packs bought by the trader are made up of the three brands. The area chart shows the numbers of each brand sold each weekend.



- How many of brand C were bought? 1 mark
- How many gift packs were sold in week 2? 1 mark
- How many of brand B were sold in week 4? 1 mark
- How many of brand A were unsold? 3 marks

Go to pp 288 for Quick Answers  
or to pp 322-4 for Worked Solutions

# Solutions

## Unit 2 Test ..... p114

- 1 Mean of 8 scores is 15.  
Sum of scores =  $8 \times 15$   
= 120  
A score of 6 is added.  
New sum is 126.  
New mean =  $126 \div 9$   
= 14

D

2  $\bar{x} = 66, s = 11$

55 is one standard deviation below the mean.  $\frac{1}{2}$  of 68% lie between 55 and 66

88 is two standard deviations above the mean.  $\frac{1}{2}$  of 95% lie between 66 and 88

Total percentage =  $34\% + 47.5\%$   
= 81.5%

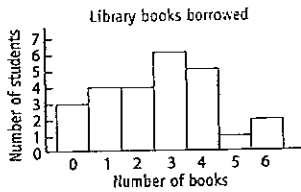
C

- 3 The correlation is weak.

A

- 4 The mode is 3.

A



- 5 Number of books borrowed  
=  $4 \times 1 + 4 \times 2 + 6 \times 3 + 5 \times 4$   
+  $1 \times 5 + 2 \times 6$   
= 67

C

6  $x = 87, \bar{x} = 63.2, s = 8.75$

$$z = \frac{x - \bar{x}}{s}$$
$$= \frac{87 - 63.2}{8.75}$$
$$= 2.72$$

D

# Solutions

## Continued next page

7 6 2 1 4 3 2 3 0 2 2 4  
3 1 3 5 21 2 3 4 2 3

The outlier is 21.

C

8 9 3 4 4 8 11 6 2  
In order: 2 3 4 4 6 8 9 11

$$\text{Median} = \frac{4+6}{2} = 5$$

B

9 Mean =  $72 - 2.4 \times 6 = 57.6$

The mean was 57.6%.

A

10 Exam results

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

X: median = 72  
Y: median = 75  
Difference =  $75 - 72 = 3$

A

11 A shows a strong negative correlation.



A

12 Ken:  $z = \frac{75-81}{7.5} = -0.8$

Barbara:  $z = \frac{62-67}{6.25} = -0.8$

Jess:  $z = \frac{68-74}{12} = -0.5$

Jess scored the best result.

C

13 Number of men =  $28 + 42 = 70$

Percentage in Australia =  $\frac{28}{70} \times 100\% = 40\%$

C

	Men	Women
Australia	28	56
Overseas	42	64

14 99.7% of scores are within 3 standard deviations of the mean.  
0.15% are more than 3 standard deviations above the mean.  
0.15% of 48 000 = 72

A

15 143 157 191 156 184 176 138  
145 152 166 171 162

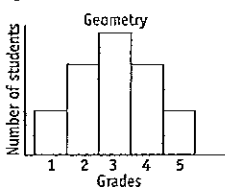
$$\begin{aligned} \bar{x} &= 161.75, s = 15.82 \\ \bar{x} + s &= 161.75 + 15.82 = 177.57 \\ \bar{x} - s &= 161.75 - 15.82 = 145.93 \end{aligned}$$

7 scores are within 1 standard deviation of the mean.

C

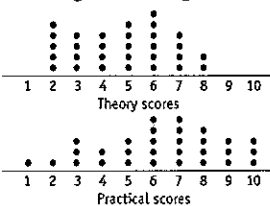
16 Geometry had the lowest standard deviation.  
[The grades in geometry are the least spread from the mean.]

B



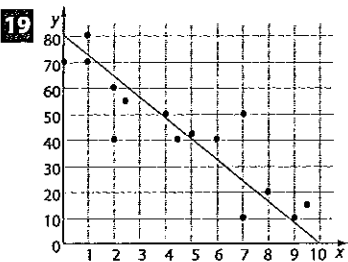
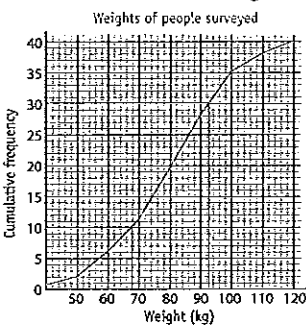
17 The mean will be higher on the practical test. [Most of the scores are higher than on the theory test.]  
The standard deviation will be higher on the practical test. [The scores are spread further in the practical test than in the theory test.]  
The mean and standard deviation were both higher in the practical test.

B



18 Interquartile range =  $(93 - 68) \text{ kg} = 25 \text{ kg}$

B



$$\begin{aligned} \text{Gradient} &= \frac{\text{vertical change}}{\text{horizontal change}} \\ &= \frac{-80}{10} \\ &= -8 \end{aligned}$$

Vertical intercept = 80

The equation is  $y = -8x + 80$

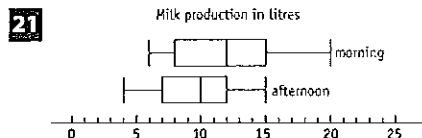
C

20 95% will lie within two standard deviations of the mean.

$$\begin{aligned} \bar{x} + 2s &= 16.5 + 2 \times 4.5 = 25.5 \\ \bar{x} - 2s &= 16.5 - 2 \times 4.5 = 7.5 \end{aligned}$$

Between 7.5 and 25.5

C



- a The median milk production in the afternoon is 10 litres. ✓  
b In the morning 8 litres is the lower quartile. Approximately 75% of cows produced more than 8 litres. ✓  
c Interquartile range =  $12 - 7 = 5$   
Interquartile range is 5 L. ✓  
d Both sets of data are fairly symmetrical, with the morning production being constantly higher than the afternoon. 25% of the cows produced more in the morning than any cow did in the afternoon. ✓✓

22 a

Score	77	89	101	113	125	137	149
z-score	-3	-2	-1	0	1	2	3

- b The median is 113. ✓  
In a normal distribution the mean and median are equal. ✓  
c Approximately 95% lie between 89 and 137. ✓  
d 34% (half of 68%) lie between 101 and 113.  
49.85% (half of 99.7%) lie between 113 and 149. ✓  
 $34\% + 49.85\% = 83.85\%$  ✓  
e 0.15% (half of 0.3%) will have scored more than 149. ✓  
 $0.15\% \text{ of } 12\,000 = 18$   
Approximately 18 students will have scored full marks. ✓  
23 a 60 ✓  
b There is a moderate positive correlation between essay marks and exam marks. ✓  
c One student scored 8 in the essay but only about 25 in the exam. ✓

$$d \quad m = \frac{\text{vertical change}}{\text{horizontal change}}$$

$$= \frac{20}{4}$$

$$= 5$$

Vertical intercept = 30  
Equation is  $y = 5x + 30$

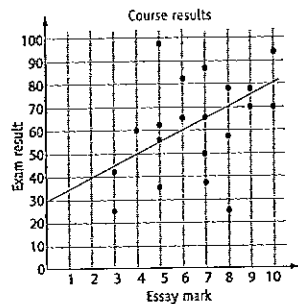
$$e \quad y = 5x + 30$$

When  $x = 7$ ,

$$y = 5 \times 7 + 30$$

$$= 65$$

Natalie's estimated mark was 65.



**24** 78 56 64 58 69 73 79 59 63  
82 57 77 63 65 56 70 62 66

a The mean = 66.5

$$b \quad \sigma_n = 8.139 \ 410 \ 298 \dots$$

$$= 8.1 \quad (1 \text{ d.p.})$$

$$c \quad z = \frac{x - \bar{x}}{s}$$

$$= \frac{73 - 66.5}{8.1}$$

$$= 0.802 \ 469 \ 135 \dots$$

$$= 0.8 \quad (1 \text{ d.p.})$$

$$d \quad z = \frac{x - \bar{x}}{s}$$

$$= \frac{68 - 60}{12.5}$$

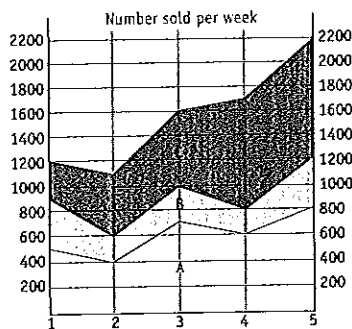
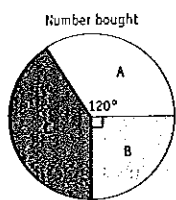
$$= 0.64$$

Nell had a better result in the first test.

**25** a Angle for C =  $360^\circ - 120^\circ - 90^\circ$   
 $= 150^\circ$

$$\text{Number of brand C} = \frac{150}{360} \times 9000$$

$$= 3750$$



b In week 2, 1100 gift packs were sold.

c In week 4, 200 of brand B gift packs were sold.

d Number of brand A bought

$$= \frac{120}{360} \times 9000$$

$$= 3000$$

$$\text{Total sold}$$

$$= 500 + 400 + 700 + 600 + 800$$

$$= 3000$$

No gift packs of brand A were unsold.