Chapter 1 Answers

Exercise 1A

1 Using a mathematical model is quicker and cheaper.
A mathematical model is a simplification which does not reflect all the aspects of the real problem.

2 Predictions based on the model are compared with observed data.
In the light of this comparison, the model is adjusted.
This process is repeated.

Chapter 2 Answers

Exercise 2A

1 **a** quantitative
b qualitative
c quantitative
d quantitative
e qualitative

2 **a** discrete
b continuous
c discrete
d continuous
e continuous
f continuous

3 **a** It is descriptive rather than rather than numerical.
b It is quantitative because it is numerical. It is discrete because its value must be an integer; you cannot have bits of a pupil.
c It is quantitative because it is numerical. It is continuous because weight can take any value in a given range.

4 **a**

Height (cm)	Frequency	Cumulative frequency
165	8	8
166	7	15
167	9	24
168	14	38
169	18	56
170	16	72

b 38 boys
c 169 cm

5 **a**

Lifetime (hours)	Frequency	Cumulative frequency
5.0–5.9	5	5
6.0–6.9	8	13
7.0–7.9	10	23
8.0–8.9	22	45
9.0–9.9	10	55
10.0–10.9	2	57

b 5.95 and 6.95 hours
c 9.45 hours

6 **a** 1.4 kg and 1.5 kg
b 1.35 kg

7 A is not true; B is true; C is true; D is true.

Exercise 2B

1 **a** 700 g
b 600 g
c 700 g
d The mean will increase; the mode will remain unchanged; the median will decrease.

2 **a** 42.7
b The mean will increase.

3 **a** 8 minutes
b 10.2 minutes
c 8.5 minutes
d The median would be best. The mean is affected by the extreme value 26.

4 increase to 24.5 litres

5 5.7 days (1 d.p.)

6 **a** 4.3 mm
b 26.5 hours
c modal rainfall 3mm, modal sunshine 15 hours
d median rainfall 2.5 mm, median sunshine 16.5 hours
e median rainfall and mean sunshine (least rainfall and highest sunshine)

7 71% (nearest percent)

Exercise 2C

1 **a**

Mark	5	6	7	8	9	10
Frequency	4	6	10	6	7	3

b 7.42
c 16
d The mean is greater.

2 5 eggs (mode 5, median 5, mean 5.44)

3 **a** 98
b 6
c 6.31
d 6
e 6

4 **a** 36
b 31
c 2
d 1
e 1.47
f the median

5 The company would use the mode (£48), since it is lower than the median (£54) and the mean (£56.10).

Exercise 2D

1 **a** 1.19
b 0.722
c The median is less than 1 and the mean is only a little above 1 – it is 1 if rounded to the nearest whole number... The hotel need not consider getting a new lift yet but should keep an eye on the situation.

2 **a** £351 to £400
b £345
c £355

3 **a** 82.3 decibels
b 16

4 Store B (mean 51 years) employs older workers than store A (mean 50 years).

5 **a** 51 mph to 60 mph
b 0.71 mph (2 d.p.) (mean 50.03 mph, median 50.74 mph)
c 9%

Exercise 2E

1 70

2 48.5

3 **a** 3.5
b **i** 7
ii 35
iii 37

4 365

5

Age (a)	Frequency (f)	Mid-point (x)	y
11–21	11	16	1.0
21–27	24	24	5.0
27–31	27	29	7.5
31–37	26	34	10.0
37–43	12	40	13.0

b 29.0

Mixed exercise 2F

1 **a** 50
b 50
c 54

2 69.2

3 **a** mean £19.57, mode £6.10, median £7.80
b The value £91.00 is wrong.

4 **a** The mean is higher than it should be.
b 34.4

5 607

6 £18 720

7 **a** group A 63.4, group B 60.2
b The method used for group A may be better.

8 **a** 7.09 km
b 7.04 km

9 **a** 25.5 minutes
b 26.6 minutes
c She spent more time each week playing computer games in the last 40 days than in the first 50 days.

10 **a** 21 to 25 hours
b 21.6 hours
c 20.6 hours
d 20.8 hours

Chapter 3 Answers

Exercise 3A

1 **a** 7
b 9
c 4
2 **a** £290
b $Q_1 = 400$, $Q_3 = 505$.
c £105
3 **a** 25, 35, 55, 65, 90, 100; total 100
b $Q_1 = 0.5$, $Q_3 = 4$.
c 3.5 hours
4 1 ($Q_1 = 9$, $Q_3 = 10$)
5 **a** 3, 9, 19, 26, 31
b 389 kg
c 480 kg
d 90.8 kg
6 **a** 1100
b 1833
c 733
7 **a** 71
b 24.6

Exercise 3B

1 **a** 8, 20, 56, 74, 89, 99
b 10
c 8
d 9
2 **a** 11, 46, 80, 96, 106, 111
b £17.10
c £28.25
d £11.15
3 £81.90
4 6.2 minutes
5 **a** 49
b 38.7 minutes
c 48.8 minutes

Exercise 3C

1 **a** 3
b 0.75
c 0.866
2 3.11 kg
3 **a** 178 cm
b 59.9 cm^2
c 7.74 cm
4 mean 5.44, standard deviation 3.25
5 **a** 25
b 4
6 **a** The mean for both routes is 14.
b Route 1 has variance 4 and standard deviation 2. Route 2 has variance 5.33 and standard deviation 2.31.
c Route 1 would be best. Although the means are the same, the standard deviation for route 1 is lower, so this route is more reliable.

Exercise 3D

1 133
2 7.35
3 **a**

Number of £'s (x)	Number of students (f)	$f\,x$	$f\,x^2$
8	14	112	896
9	8	72	648
10	28	280	2800
11	15	165	1815
12	20	240	2880
Totals	85	869	9039

b 1.82
c £1.35
4 **a**

Number of days absent (x)	Number of students (f)	$f\,x$	$f\,x^2$
0	12	0	0
1	20	20	20
2	10	20	40
3	7	21	63
4	5	20	80

b 1.51
c 1.23

5 **a**

Lifetime in hours	Number of parts	Mid-point (x)	f x	f x^2
5 < h = 10	5	7.5	37.5	281.25
10 < h = 15	14	12.5	175.0	2187.50
15 < h = 20	23	15.5	402.5	7043.75
20 < h = 25	6	22.5	135.0	3037.50
25 < h = 30	2	27.5	55.0	1512.50

b variance 22.0, standard deviation 4.69 hours

6 variance 21.25, standard deviation 4.61

6 **a** mean 15.8, standard deviation 2.06
b The mean wing span will decrease.

7 **a** 98.75
b 104
c 5.58
d 4.47

Exercise 3E

1 **a** 5.08
b **i** 5.08
ii 5.08
iii 5.08

2 **i** 70.7
ii 70.7
iii 70.7

3 **a** 0.28
b 0.675
c 2.37
d 6.5

4 2.34

5 1.76 hours

6 22.9

7 416

Mixed exercise 3F

1 **a** 6
b 3
c 9
d 6

2 37.5

3 **a** 20.5
b 34.7
c 14.2

4 15.5 m

5 **a** 40.9
b 54
c 13.1
d 10.1

Chapter 4 Answers

Exercise 4A

1

	Key: 2 | 3 means 23 DVDs.	
0	6 9	(2)
1	2 2 2 5 5 5 7 8 9	(9)
2	0 2 3 5 5 5 6 6 7 7 9 9	(12)
3	2 2 4 4 5	(5)
4	2 5	(2)

a 25

b 15

c 29

2 **a** 24
b 49
c 8
d 3
e 37
f 34
g 21
h 37

3 **a** 41
b 32
c 47
d 15
e 47

4 **a**

Boys			Girls		
(2)	9 8	2	4 6 8		(3)
(3)	4 2 2	3	2 3 4 4 9		(5)
(5)	8 7 5 5 4	4	5 6 7		(3)
(5)	7 6 6 4 4	5	2 4		(2)
(1)	0	6			

Key: 2 | 3 | 4 means 32 boys and 34 girls.

b The girls gained lower marks than the boys.

5 **a** 17 males, 15 females
b £48
c Males earned more in general.

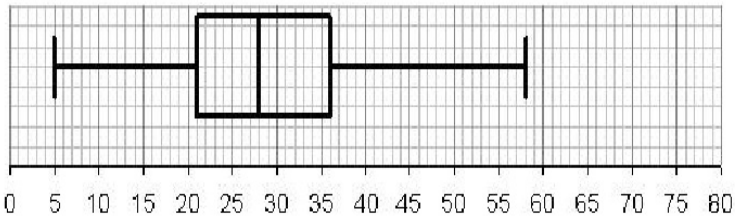
Exercise 4B

1 **a** 7 is an outlier.
b 88 is not an outlier.
c 105 is an outlier.

2 **a** no outliers
b 170 g and 440 g
c 760 g

Exercise 4C

1



2 **a** 47, 32
b 38
c 15
d 64

1 **a** 45
b lower quartile
c Boys have a lower median and a larger spread.
or Girls have a higher median and a smaller spread.

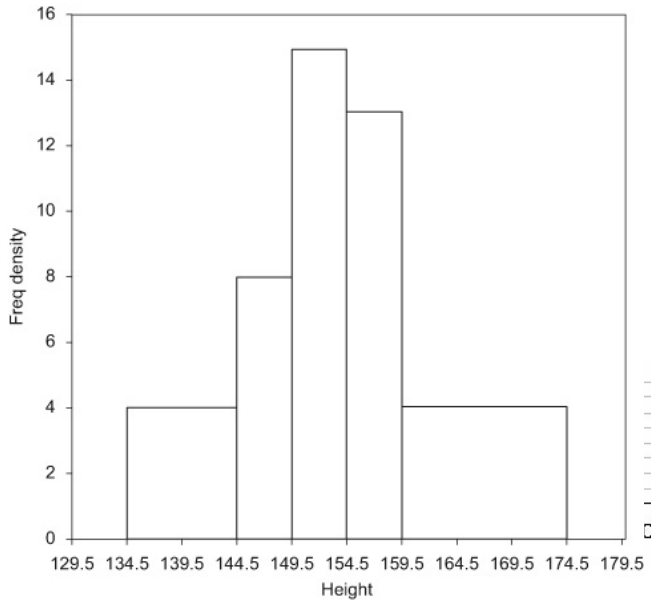
2 **a** The male turtles have a higher median weight, a greater interquartile range and a greater total range.
b It is more likely to have been female. Very few of the male turtles weighed this little, but more than a quarter of the female turtles weighed more than this.
c 500 g

Exercise 4E

1 **a**

Height (cm)	Frequency	Class width	Frequency density
135–144	40	10	4
145–149	40	5	8
150–154	75	5	15
155–159	65	5	13
160–174	60	15	4

b



2 **a** The quantity (time) is continuous.
b 150
c 369
d 699

3 **a** 114
b 90
c 24

4 **a** The quantity (distance) is continuous.
b 620
c 150
d 190
e 130

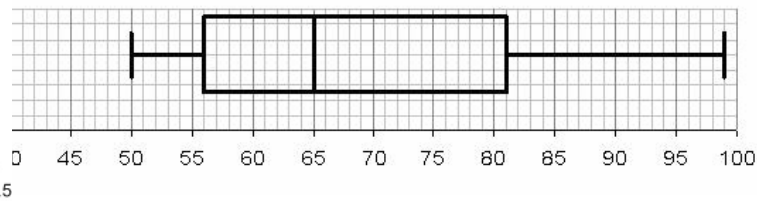
5 **a** The quantity (weight) is continuous.
b The area of the bar is proportional to the frequency.
c 0.125
d 168
e 88

Exercise 4F

1 negative skew

2 **a** mean 31.1 minutes, variance 78.05
b median 29.7 minutes; quartiles 25.8 minutes, 34.8 minutes
c 0.0853 (positive skew)
d They will use the median and quartiles because of the skew.

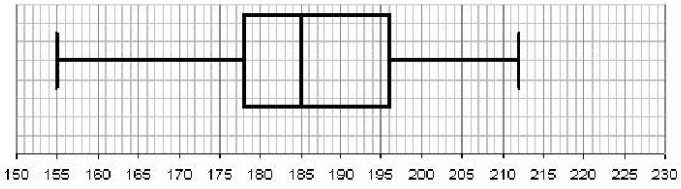
3 **a** 64 mm
b median 65 mm; quartiles 56 mm, 81 mm
c



d & f The mean is greater than the median, so the data is positively skewed.
e mean 68.7 mm, standard deviation 13.7 mm
g various answers

Mixed exercise 4G

1 **a** $Q_1 = 178$, $Q_2 = 185$, $Q_3 = 196$.
b 226
c



d positive skew

2 **a** 22
b $X = 11$, $Y = 27$, $Z = 22$.
c Strand Road has more pedal cycles, since its median is higher.

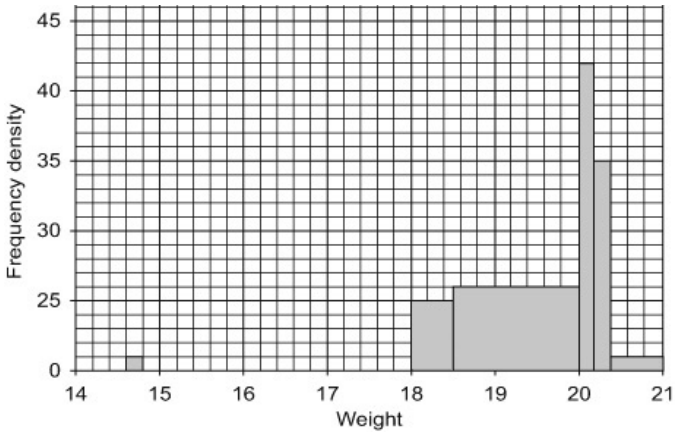
3 **a** It is true. 60 is the median for shop B.
b It is true. 40 is the lower quartile for shop A.
c Shop A has a greater interquartile range and a greater total range than shop B. Shop B has a higher median.
d Shop B is more consistent.

4 **a** 45 minutes
b 60 minutes
c This represents an outlier.
d Irt has a higher median than Esk. The interquartile ranges were about the same.
e Esk positive skew, Irt symmetric
f Esk had the fastest runners.

5 **a** 26
b 17

6 **a** 2.6 cm
b 0.28 cm

7 **a**

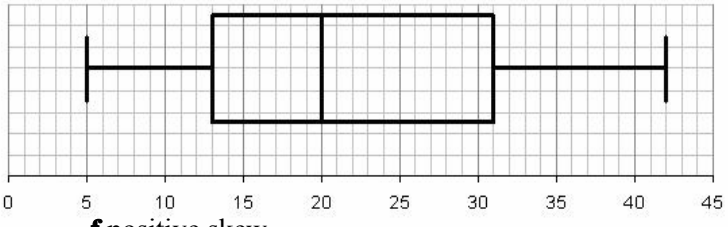


b mean 19.8 kg, s.d. 0.963 kg
c 20.1 kg
d -1.06
e negative skew

8 **a** 22.3
b

	Key: 1 | 3 means 13 bags.	
0	5	(1)
1	0 1 3 5 7	(5)
2	0 0 5	(3)
3	0 1 3	(3)
4	0 2	(2)

c median 20; quartiles 13, 31
d no outliers
e



f positive skew

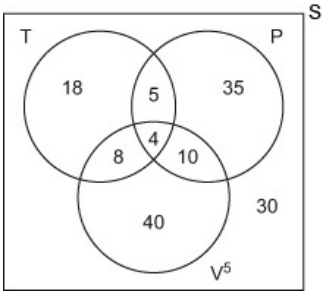
Chapter 5 Answers

Exercise 5A

1 0.5
2 0.5
3 0.25
4 0.125
5 0.0833
6 0.167

Exercise 5B

1 **a** 0.0769
b 0.25
c 0.0192
d 0.308
e 0.75
f 0.231
2 **a** 0.56
b 0.24
c 0.32
d 0.04
3 **a** 0.6
b 0.1
c 0.4
5 **a** 0.12
b 0.08
c 0.08
d 0.432
6 **a**



b 0.324
c 0.375
d 0.255
e 0.371

Exercise 5C

1 **a** 0.6
b 0.8
c 0.4
d 0.9
2 **a** 0.3
b 0.6
c 0.8
d 0.9
3 **a** 0.25
b 0.5
c 0.65
d 0.1
4 **a** 0.15
b 0.45
c 0.55
d 0.25
e 0.3
5 0.1
6 **a** 0.17
b 0.18
c 0.55
7 **a** 0.3
b 0.3

Exercise 5D

1 0.0769
2 **a** 0.333
b 0.667
3 **a** 0.182
b 0.727
4 **a** 0.7
b 0.667
c 0.8
d 0.4
5 **a** 0.5
b 0.3
c 0.3
6 **a** 0.3
b 0.35
c 0.4

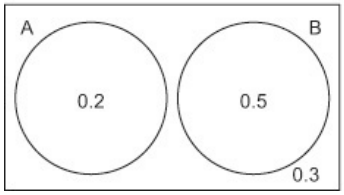
7 **a** 0.0833
b 0.15
c 0.233
d 0.357
e 0.643
f 0.783

Exercise 5E

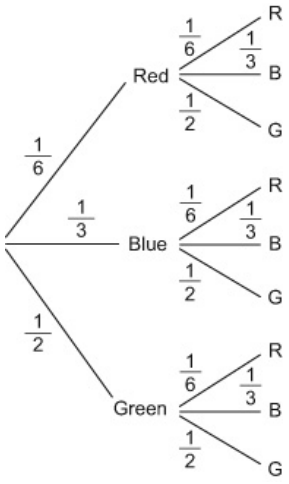
1 **a** 0.625
b ½
c 0.167
d 0.555
2 **a** 0.163
b 0.507
3 0.36
4 **a** 0.25
b 0.333
5 If the contestant sticks, their probability of winning is 1/3. If they switch, the probability of winning is 2/3. So they should switch. (This answer assumes that the host knows where the sports car is.)

Exercise 5F

1 **a**



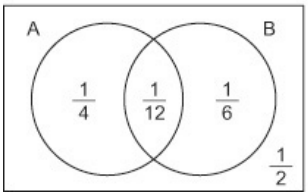
b 0.7
c 0.3
2 **a** 0.05
b 0.2
c 0.6
3 **a** mutually exclusive
b 0.6
c 0.4
4 **a** *various*
b *various*
5 **a** 0.391
b 0.625
6 **a** 0.0278
b 0.0217
c 0.290
7 **a**



b 0.389
c 0.611
8 **a** 0.0156
b 0.911
c 0.0675
d 0.0203

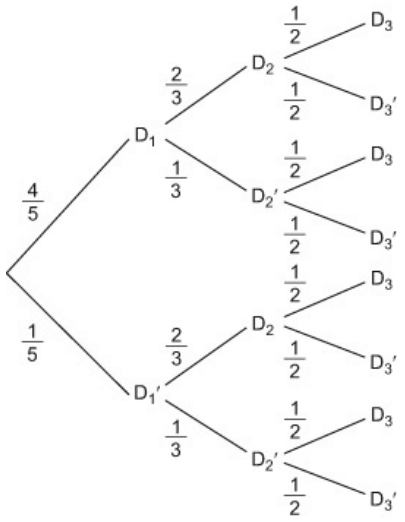
Mixed exercise 5G

1 **a** *various*
b



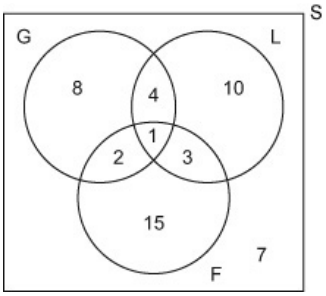
c 0.333

2 **a**



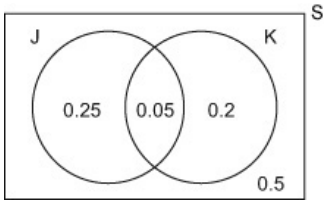
b 0.267
c 0.233

3 **a**



b 0.3
c 0.14
d 0.25

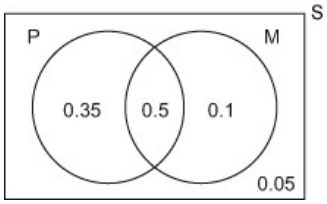
4 **a**



b 0.3
c 0.25
d 0.2

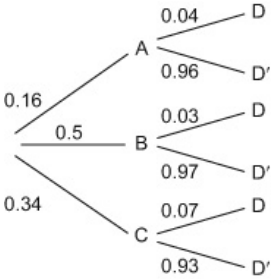
5 **a** 0.123
b 0.231

6 **a** 0.5
b



c 0.895

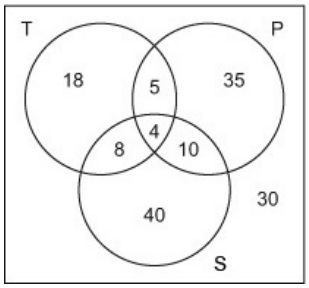
7 **a**



b 0.015
c 0.0452
d 0.332

8 **a** 0.2
b 0.5
c 0.245
d 0.571

9 **a**

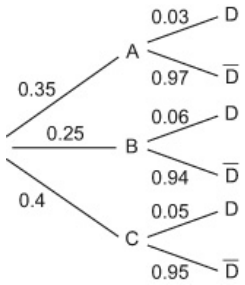


b 0.2
c 0.82
d 0.430
e 0.169

10 **a** 0.32
b 0.46
c 0.22
d 0.2016

Review Exercise 1

1 **a**



b i 0.0105
ii 0.0455
c 0.440

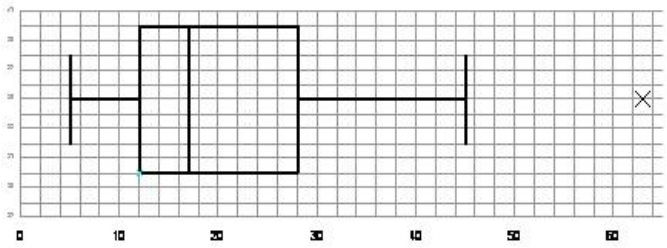
2 **a** positive skew
b median 26.7 miles
c mean 29.6 miles, standard deviation 16.6 miles
d 0.520
e yes; 0.520 > 0
f The median would be best, since the data is skewed.
g The distribution is symmetric (or has zero skew).

3 **a** Time is a continuous quantity.
b Area is proportional to frequency.
c *various*
d 30

4 **a** *any two of the following:*
Statistical models simplify a real world problem.
They are cheaper and quicker than an experiment.
They are easier to modify than an experiment.
They improve understanding of problems in the real world.
They enable us to predict outcomes in the real world.
b 3. The model is used to make predictions.
4. Experimental data is collected.
7. The model is refined.

5 **a** Distance is a continuous quantity.
b 0.8, 3.8, 5.3, 3.7, 0.75, 0.1
c $Q_2 = 58.8$, $Q_1 = 52.5$, $Q_3 = 67.1$.
d 62.5 km
e 0.137, positive skew
f The mean is greater than the median.

6 **a**

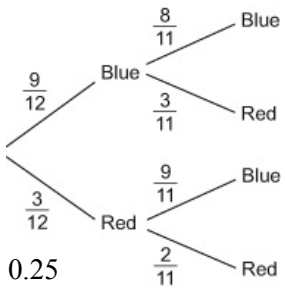


b The distribution is positively skewed, since $Q_2 - Q_1 < Q_3 - Q_2$
c Most of the delays are so small that passengers should find them acceptable.

7 **a** 0.338
b 0.46
c 0.743
d 0.218

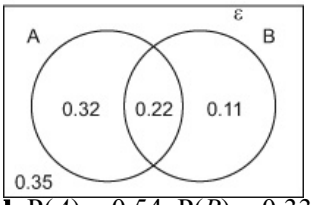
8 **a** 56
b $Q_1 = 35$, $Q_2 = 52$, $Q_3 = 60$.
c mean = 49.4, s.d. = 14.6
d -0.448
e The mean (49.4) is less than the median (52), which is less than the mode (56).

9 **a**



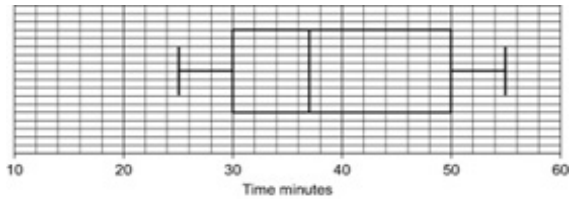
b 0.25
c 0.182

10 a



b P(A) = 0.54, P(B) = 0.33.
c 0.478
d They are not independent.

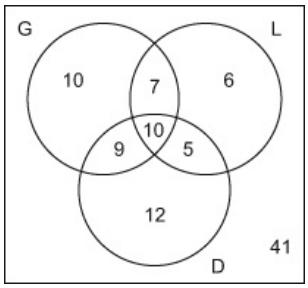
11 a maximum, minimum, median, quartiles, outliers
b i 37 minutes **ii** upper quartile, third quartile, 75 percentile
c outliers – values that are much greater than or much less than the other values and need to be treated with caution
d



e The children from school A generally took less time than those from school B.
The median for A is less than the median for B.
A has outliers, but B does not.
Both have positive skew.
The interquartile range for A is less than the interquartile range for B.
The total range for A is greater than the total range for B.

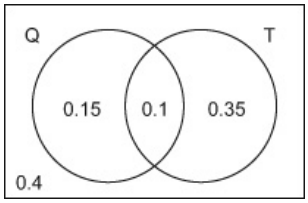
12 a 0.0370
b 19.3 minutes
c 24.8 minutes
d Their conversations took much longer during the final 25 weeks.

13 a



b 0.1
c 0.41
d 0.21
e 0.667

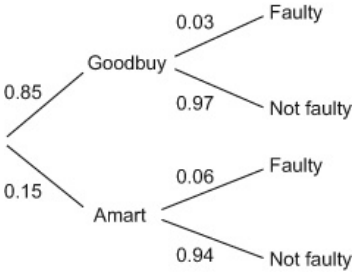
14 a 0.1
b



c 0.25

15 a 35, 15
b 40
c 18.9 minutes
d 7.26 minutes
e median 18 minutes; quartiles 13.75 minutes, 23 minutes
f 0.376, positive skew

16 a



b 0.9655

17 mean 240, standard deviation 14

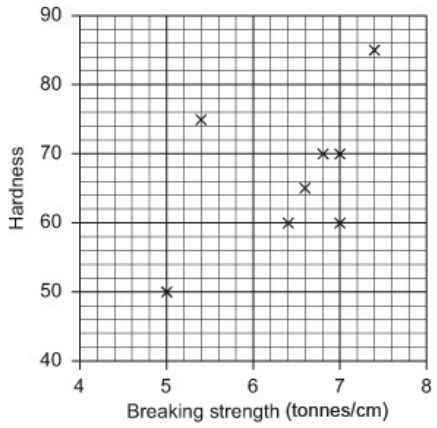
Chapter 6 Answers

Exercise 6A

1 **a i** no correlation
ii negative correlation
iii positive correlation
b i There is no correlation between height and intelligence.
ii As age increases, price decreases.
iii As length increases, breadth increases.

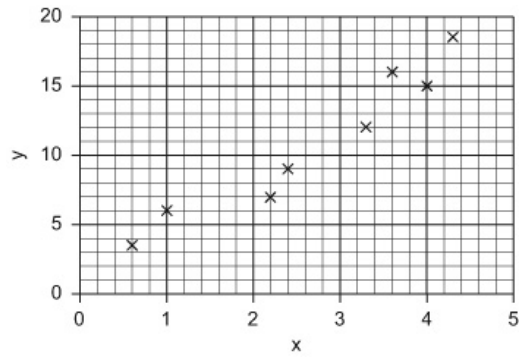
2 **a** positive correlation
b The longer the treatment, the greater the loss of weight.

3 **a**



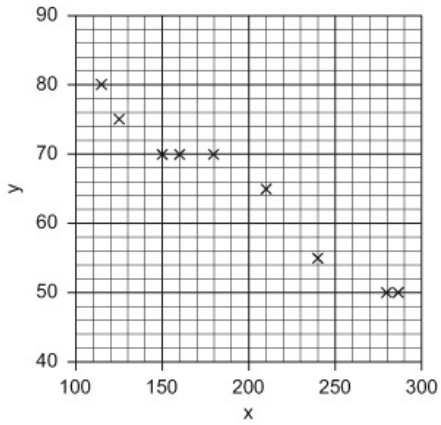
b There is weak positive correlation. There is some reason to believe that, as breaking strength increases, hardness increases.

4 **a**



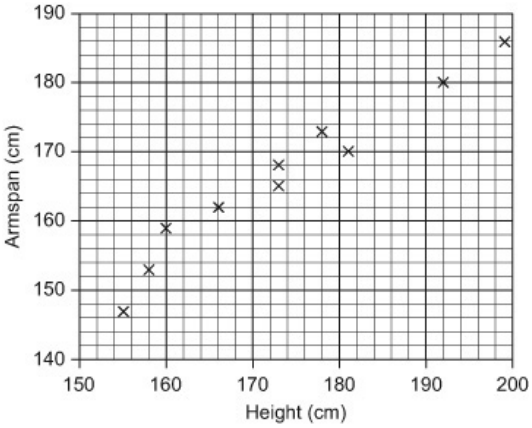
positive correlation

b



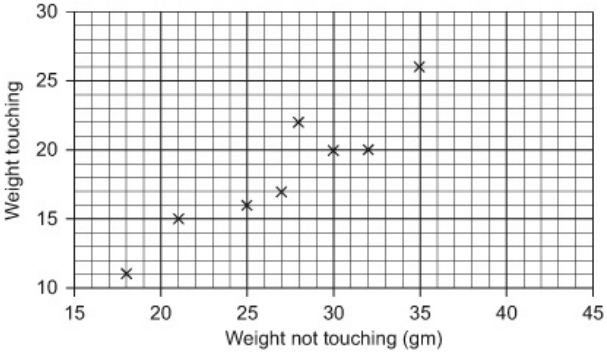
negative correlation

5 **a**



b There is positive correlation. As height increases, arm-span increases.

6 **a**



b There is positive correlation. If a student guessed a greater weight before touching the bag, they were more likely to guess a greater weight after touching it.

Exercise 6B

1 1.775

2 7.90

3 -14

4 0.985

5 0.202

6 **a** 9.71
b 0.968
c There is positive correlation. The greater the age, the taller the person.

7 **a** $S_{ll} = 30.3$, $S_{tt} = 25.1$, $S_{lt} = 25.35$.
b 0.919

8 **a** 0.866
b There is positive correlation. The higher the IQ, the higher the mark in the intelligence test.

9 **a** $S_{xx} = 82.5$, $S_{yy} = 32.9$, $S_{xy} = -44.5$.
b –0.854
c There is negative correlation. The relatively older young people took less time to reach the required level.

Exercise 6C

1 **a** iii (0)
b i (-0.96)

2 **a** i (-1)
b iii (0)

3 There is strong positive correlation. The taller the father is, the taller his son will be.

4 **a** ii
b iv
c iii
d i

5 There is strong positive correlation between x and y. As x increases, y increases.
There is strong negative correlation between s and t. As s increases, t decreases.

6 This is not sensible. There is no way in which one could be directly caused by the other.

7 This is not sensible. It is more likely that the taller pupils are older.

Exercise 6D

1 **a** x - 2000, $y/3$
b $s/100$, no change to t

2 0.973

3 0.974

4 **a** $S_{pp} = 10$, $S_{tt} = 5.2$, $S_{pt} = 7$.
b 0.971
c 0.971

5 **a**

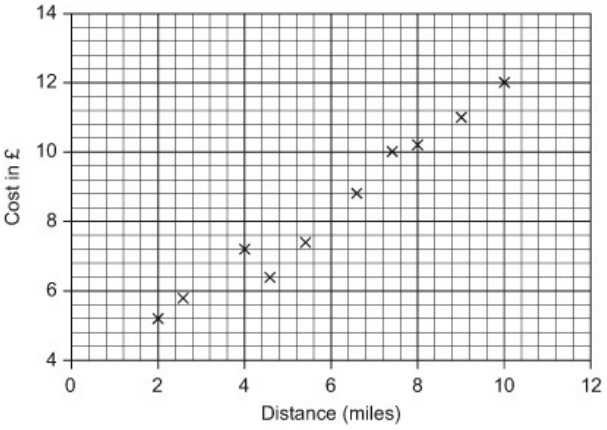
x	15	37	5	0	45	27	20
y	30	13	34	43	20	14	0

b $S_{xx} = 1.59$, $S_{yy} = 62$, $S_{xy} = 8.1$.
c 0.816
d 0.816
e The greater the mass of a woodmouse, the longer its tail.

6 **a** $S_{xx} = 1601$, $S_{yy} = 1282$, $S_{xy} = -899$.
b -0.627
c The shopkeeper is wrong. There is negative correlation. Sweet sales actually decrease as newspaper sales increase.

Mixed exercise 6E

1 **a**



b There is correlation positive. The further the taxi travels, the more it costs.

2 **a** **i** shows positive correlation.
ii shows negative correlation.
iii shows no correlation.
b **i** The older a snake is, the longer it is.
ii The higher the unemployment, the lower the drop in wages.
iii – There is no correlation between the age and height of men.

3 **a** ii (-0.12)
b i (0.87)
c iii (-0.81)

4 As a person's age increases, their score on the memory test decreases.

5 **a** -0.147
b -0.147
c This is weak negative correlation. There isjust a little evidence to suggest that students in the group who are good at science are also good at art.

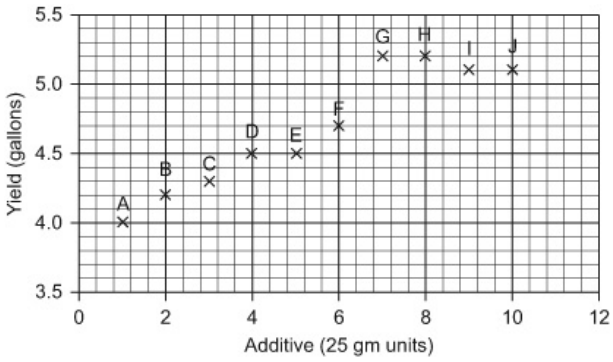
6 **a** $S_{jj} = 4413$, $S_{pp} = 5145$, $S_{jp} = 3972$.
b 0.834
c There is strong positive correlation, so Nimer is correct.

7 **a**

x	20	40	30	52	15	5	80	40	5	0
y	7	10	9	10.5	6	5	10	9	6	3

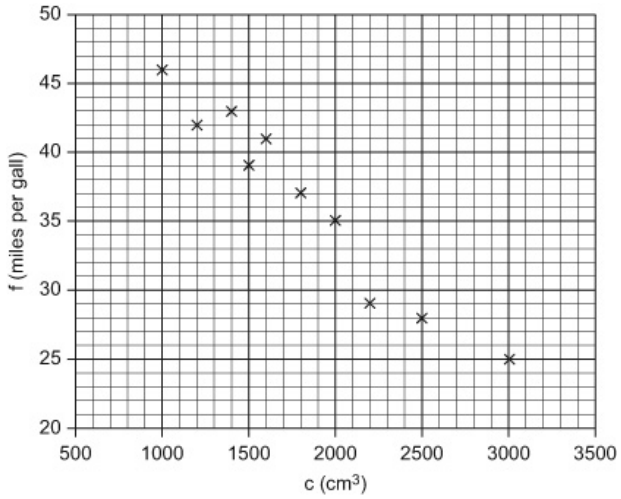
b $S_{xx} = 5642$, $S_{yy} = 57.2$, $S_{xy} = 494$.
c 0.870
d 0.870
e This is a strong positive correlation. As v increases, m increases.

8 **a**



b There is some positive correlation. Using the additive increases milk yield.
c Each cow should be given 7 units. The yield levels off at this point.
d The yield stops rising after the 7th cow (G).
e 0.952
f It would be less than 0.952. The yield of the last 3 cows is no greater than that of the 7th cow.

9 a



b There is negative correlation. As engine size increases, the number of miles per gallon decreases.
c $S_{cf} = -38\,200$.
d $c/100$ *or* $(c - 1000)/100$ and $f - 25$.

10 a $S_{xx} = 91.5$, $S_{yy} = 38.9$, $S_{xy} = 32.3$.
b 0.541
c 0.541
d There is positive correlation. As age increases, blood pressure increases.

Chapter 7 Answers

Exercise 7A

1 The number of operating theatres is the independent variable.
The number of operations is the dependent variable.

2 The number of suitable habitats is the independent variable.
The number of species is the dependent variable.

3 $a = -3$, $b = 6$.

4 $y = -14 + 5.5x$

5 $y = 2x$

6 **a** $S_{xx} = 5$, $S_{xy} = 20$.
b $y = 2 + 4x$

7 **a** $S_{xx} = 40.8$, $S_{xy} = 69.6$.
b $y = -0.294 + 1.71x$

8 $g = 1.50 + 1.44h$

9 **a** $S_{nn} = 6486$, $S_{np} = 6344$.
b $p = 20.9 + 0.978n$

10 **a** $S_{xx} = 10$, $S_{xy} = 14.5$.
b $y = -0.07 + 1.45x$

Exercise 7B

1 $y = 6 - x$

2 $s = 88 + p$

3 $y = 32 - 5.33x$

4 $t = 9 + 3s$

5 **a** $y = 3.5 + 0.5x$
b $d = 35 + 2.5c$

6 **a** $S_{xy} = 162$, $S_{xx} = 191$; $y = 7.87 + 0.85x$
b $y = 22.35 + 2.125h$

Exercise 7C

1 6

2 384 g

3 **a** Extrapolation means using the regression line to estimate outside the range of the data. It can be unreliable.
b Interpolation means using the regression line to estimate within the range of the data. It is usually reasonably reliable.

4 **a** £6.00. This is reliable since 7 years is within the range of the data used.
b 3 years is outside the range of the data.

5 This is not a sensible estimate since 30 minutes is a long way outside the range of the data.

6 **a** 3845. This is reliable since £2650 is within the range of the data.
b 9730. This is unreliable since £8000 is outside the range of the data.
c 1100 extra books are sold for each £1000 spent on advertising.
d It suggests that 930 books would be sold if no money were spent on advertising. This is reasonably reliable since it is only just outside the range of the data.

7 **a** 283
b If dexterity increases by 1 unit, production increases by 57 units.
c i The estimate would be reliable. 2 is outside the range of the data, but only just.
ii The estimate would be unreliable. 14 is a long way outside the range of the data.

8 The equation of the regression line would change.

Mixed exercise 7D

1 **a** 2.45 mm
b 4.52 mm
c The answer to part a is reliable since 300^0C is within the range of the data. The answer to part b is unreliable since 530^0C is outside the range of the data.

2 **a** $t = 1.96 + 0.95s$.
b 49.4

3 **a** $S_{xx} = 6.43$, $S_{xy} = 11.7$.
b $y = 0.554 + 1.82x$
c 6.01 cm

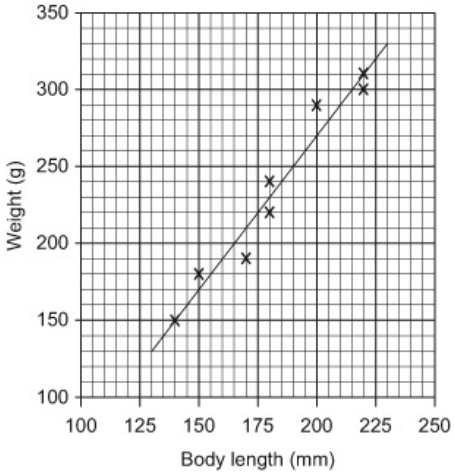
4 **a** $S_{xx} = 16\,350$, $S_{xy} = 210\,330$
b $y = 225 + 12.9x$
c 1510
d 255
e This answer is unreliable since a gross national product of 3500 is long way outside the range of the data.

5 **a** $y = 0.343 + 0.499x$
b $t = 2.34 + 0.224m$
c 4.58 cm

6 **a** $S_{xx} = 90.9$, $S_{xy} = 190$.
b $y = -1.82 + 2.09x$
c $p = 25.5 + 2.09r$
d 71.4
e This answer is reliable since 22 breaths per minute is within the range of the data.

7 **a** 0.79 kg is the average amount of food consumed in 1 week by 1 hen.
b 23.9 kg
c £476

8 **a & e**



b There appears to be a linear relationship between body length and body weight.
c $w = -12.7 + 1.98l$
d $y = -127 + 1.98x$
f 289 g. This is reliable since 210 mm is within the range of the data.
g Water voles B and C were probably removed from the river since they are both underweight. Water vole A was probably left in the river since it is slightly overweight.

9 **a** $S_{xy} = 78$, $S_{xx} = 148$.
b $y = 7.31 + 0.527x$
c $w = 816 + 211n$.
d 5036 kg
e 100 items is a long way outside the range of the data.

Chapter 8 Answers

Exercise 8A

1 **a** This is not a discrete random variable, since time is a continuous quantity.
b This is not a discrete random variable, since it is always 7 and thus does not vary
c This is a discrete random variable, since it is always a whole number and it does vary.

2 0, 1, 2, 3, 4

3 **a** {(2, 2), (2, 3), (3, 2), (3, 3)}
b

x	4	5	6
P($X = x$)	0.25	0.5	0.25

c $P(X = 4) = 0.25$, $P(X = 5) = 0.5$, $P(x = 6) = 0.25$.

4 0.0833

5 $k + 2k + 3k + 4k = 1$, so $10k = 1$, so $k = 0.1$.

6

x	1	2	3	4	5
P($X = x$)	0	0.1	0.2	0.3	0.4

7 **a** 0.125
b

x	1	2	3	4
P($X = x$)	0.125	0.125	0.325	0.325

8 **a** 0.3
b

x	-2	-1	0	1	2
P($X = x$)	0.1	0.1	0.3	0.3	0.2

9 0.25

Exercise 8B

1 **a** 0.5
b 0.2
c 0.6

2 **a** 0.625
b 0.375

3 **a**

x	1	2	3	4	5	6
F(x)	0.1	0.2	0.35	0.60	0.9	1.0

b 0.9
c 0.2

4 **a**

x	1	2	3	4	5	6
P(X = x)	0.1	0.1	0.25	0.05	0.4	0.1

b 0.5
c 0.4

5 **a** 0.0556
b

x	P($X = x$)
1	0.0556
2	0.0556
3	0.1667
4	0.1667
5	0.2777
6	0.2777

c 0.389
d 0.444
e 0.0556

6 **a** 0.3
b

x	-2	-1	0	1	2
P(X = x)	0.1	0.1	0.25	0.25	0.3

c 0.45

7 **a** 0.833
b 0.167
c

x	1	2	3	4	5
P(X = x)	0.333	0.167	0.167	0.167	0.167

8 **a** 1
b

x	1	2	3
P($X = x$)	0.25	0.3125	0.4375

Exercise 8C

1 **a** $E(X) = 4.6$, $E(X^2) = 26$.
b $E(X) = 2.8$, $E(X^2) = 9$.

2 $E(X) = 4$, $E(X^2) = 18.2$.

3 **a**

x	2	3	6
P(X = x)	0.5	0.333	0.167

b E(X) = 3, E(X^2) = 11.
c no

4 **a**

Number of heads (h)	0	1	2
P(H = h)	0.25	0.5	0.25

b 0 heads 12.5 times, 1 head 25 times, 2 heads 12.5 times
c The coins may be biased. There were more times with 2 heads and fewer times with 0 heads than expected.

5 $a = 0.3$, $b = 0.3$.

6 100

Answers 8D

1 **a** 1
b 2

2 **a** E(X) = 3.83, Var(X) = 0.472.
b E(X) = 0, Var(X) = 0.5.
c E(X) = -0.5, Var(X) = 2.25.

3 E(Y) = 3.5, Var(X) = 2.917.

4 **a**

x	P(X = x)
2	0.0278
3	0.0556
4	0.0833
5	0.1111
6	0.1389
7	0.1667
8	0.1389
9	0.1111
10	0.0833
11	0.0556
12	0.0278

b 7
c 5.833

5 **a**

d	0	1	2	3
P(D = d)	0.25	0.375	0.25	0.125

P(D = 3) = 0.125.
b 1.25
c 0.9375

6 **a** P(T = 1) = P(head) = 0.5,
P(T = 2) = P(tail, head) = 0.5 × 0.5 = 0.25,
P(T = 3) = 1 - P(T = 1) - P(T = 2) = 0.25.
b E(T) = 1.75, Var(T) = 0.687.

7 **a** 2
b $a = 0.375$, $b = 0.25$.

Answers 8E

1 **a** 8
b 40

2 **a** 6
b 7
c 1
d 0
e 54
f 54
g 6

3 **a** 7
b 5
c 36
d 9

4 **a** 4μ
b $2\mu + 2$
c $2\mu - 2$
d $4\sigma^2$
e $4\sigma^2$

5 **a** 7
b -4
c 81
d 81
e 13
f 12

6 E(S) = 64, Var(S) = 225.

7 **a**

x	1	2	3
P(X = x)	0.25	0.375	0.375

b 2.125
c 0.609
d 5.25
e 5.48

8 **a** 0.2
b 0.76
c 1.07
d 0.0844

Exercise 8F

1 E(X) = 3, Var(X) = 2.

2 **a** 4
b 4

3 **a** E(X) = 3.5, Var(X) = 2.92.
b 0.667

4 **a** 0.3
b E(X) = 11, Var(X) = 33.

5 **a** 0.2
b E(X) = 10, Var(X) = 33.

6 A discrete uniform distribution is not a good model. The game depends on the skill of the player. The points are likely to cluster around the middle.

7 **a** a discrete uniform distribution
b 4.5
c 5.25
d The expected winnings are less than the 5p stake.

Mixed Exercise 8G

1 **a**

x	P(X = x)
1	0.0476
2	0.0952
3	0.1429
4	0.1905
5	0.2381
6	0.2857

b 0.571
c 4.33
d 2.22
e 8.89

2 **a** 0.2
b 0.7
c 0.6
d 3.6
e 8.04

3 **a** 0.3
b
E(X) = 0 × 0.2 + 1 × 0.3 + 2 × 0.5 = 1.3.
c 0.61
d 0.5

4 **a** $k + 0 + k + 2k = 1$,
so $4k = 1$, so $k = 0.25$.
b E(X) = 2.
E(X^2) = $0^2 \times 0.25 + 1^2 \times 0 + 2^2 \times 0.25 + 3^2 \times 0.5$
= 1 + 4.5 = 5.5.
c 6

5 **a** 0.125
b 0.75
c 1.125
d 2.375
e 0.859

6 **a** discrete uniform distribution
b *any discrete distribution where all the probabilities are the same*
c 2
d 2

7 **a** $p + q = 0.5$, $2p + 3q = 1.3$.
b $p = 0.2$, $q = 0.3$.
c 1.29
d 5.16

8 **a** 0.111
b 3.44
c Var(X) = E(X^2) - E(X)^2
˜ $13.88889 - 3.44444^2$
˜ 2.02.
d 8.1 (2 s.f.)

9 **a** E(X) = 3.5.
Var(X) = E(X^2) - E(X)^2
= 91/6 - $(7/2)^2$ = 35/12.
b 6
c 11.7

10 **a**

x	1	2	3	4
P(X = x)	0.0769	0.1923	0.3077	0.4231

b 0.731
c 3.077
d $\text{Var}(X) = \text{E}(X^2) - \text{E}(X)^2$
$\approx 10.385 - 1.077^2 \approx 0.92$
e 8.28

Chapter 9 Answers

Exercise 9A

1 **a** 0.9830 **b** 0.9131 **c** 0.2005 **d** 0.3520
2 **a** 0.1056 **b** 0.9535 **c** 0.0643 **d** 0.9992
3 **a** 0.9875 **b** 0.4222 **c** 0.4893 **d** 0.0516
4 **a** 0.0902 **b** 0.9438 **c** 0.1823 **d** 0.8836

Exercise 9B

1 **a** 1.33 **b** 1.86 **c** -0.42 **d** -0.49
2 **a** 2.50 **b** 0.22 **c** -0.71 **d** 0.8416
3 **a** 1.0364 **b** -1.6449 **c** 1.22 **d** 3.0902
4 **a** 1.06 **b** 2.55 **c** 0.81 **d** 1.35
5 **a** 0.2533 **b** 1.0364 **c** 1.2816 **d** 0.5244

Exercise 9C

1 **a** 0.9332 **b** 0.9772
2 **a** 0.0475 **b** 0.2514
3 **a** 0.1587 **b** 0.4985
4 **a** 0.264 **b** 0.171
5 **a** 0.961 *or* 0.962 **b** 22.4
6 32.6
7 18.1
8 **a** 19.1 **b** 18.3 **c** 0.0915
9 **a** 70.6 **b** 80.8 **c** 0.075
10 **a** 81.0 **b** 80.6 **c** 0.0364

Exercise 9D

1 11.5
2 3.87
3 31.6
4 25
5 $\mu = 13.1$, $\sigma = 4.32$.
6 $\mu = 28.3$, $\sigma = 2.59$.
7 $\mu = 12$, $\sigma = 3.56$.
8 $\mu = 35$, $\sigma = 14.8$ *or* $\sigma = 14.9$.
9 4.75
10 $\sigma = 1.99$, $a = 2.18$.

Mixed exercise 9E

1 **a** 0.0401 **b** 188 cm
2 **a** 12.7% *or* 12.8% **b** 51.1% *or* 51.2%
3 **a** 0.0668 **b** 0.052
4 **a** 3.65 **b** 0.1357 **c** 32.5
5 **a** 8.60 ml **b** 0.123 **c** 109 ml
6 $\mu = 30$, $\sigma = 14.8$ *or* $\sigma = 14.9$
7 mean 3.76 cm, standard deviation 10.2 cm
8 **a** 0.3085 **b** 0.370 *or* 0.371
c The first score was better, since fewer of the students got this score or more.
9 **a** 4.25 *or* 4.26 **b** 0.050
10 **a** 8.54 minutes **b** 0.176
11 mean 6.12 mm, standard deviation 0.398 mm
12 **a** 0.8413 **b** 0.111
13 **a** 0.2119 **b** 28.2

Review Exercise 2

1 **a** £17
b $S_{tm} = 1191.8$, $S_{tt} = 983.6$, $S_{mm} = 1728.9$.
c 0.914
d 0.914. Linear coding does not affect the correlation coefficient.
e 0.914 suggests a relationship between the time spent shopping and the money spent. 0.178 suggests that there was no such relationship.
f *various*

2 **a**

x	$P(X = x)$
1	0.0278
2	0.0833
3	0.1389
4	0.1944
5	0.2500
6	0.3056

b 0.583
c 4.47
e 17.7

3 **a** 0.2743
b 12

4 Diagram A corresponds to -0.79, since there is negative correlation Diagram B. corresponds to 0.08, since there is no significant correlation.
Diagram C corresponds to 0.68, since there is positive correlation.

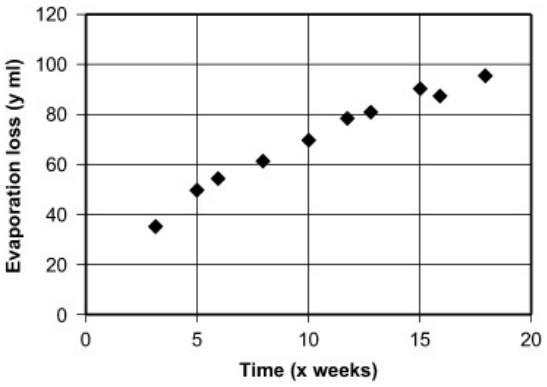
5 **a** $y = -0.425 + 0.395x$
b $f = 0.735 + 0.395m$
c 93.6 litres

6 **a** 0.0588
b 3.76
c 1.47
d 13.3

7 **a** 0.076 *or* 0.077
b 0.639 *or* 0.640
c 153.2

8 **a** $p + q = 0.4$, $2p + 4q = 1.3$.
b $p = 0.15$, $q = 0.25$.
c 1.75
d 7.00

9 **a**



b The points lie close to a straight line.
c $a = 29.02$, $b = 3.90$.
d 3.90 ml of the chemicals evaporate each week.
e i 103 ml
ii 166 ml
f i This estimate is reasonably reliable, since it is just outside the range of the data.
ii This estimate is unreliable, since it is far outside the range of the data.

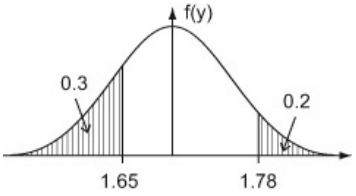
10 **a** A statistical model simplifies a real world problem.
It improves the understanding of a real world problem.
It is quicker and cheaper than an experiment or a survey.
It can be used predict possible future outcomes.
It is easy to refine a statistical model.
b i *various*
ii *various*

11 **a** 0.0618
b 0.9545
c 0.00281
d This is a bad assumption.

12 **a** $S_{xy} = 71.47$, $S_{xx} = 1760$
b $y = 0.324 + 0.0406x$
c 2461.95 mm
d $l = 2460.324 + 0.0406t$
e 2463.98 mm
f This estimate is unreliable since it is outside the range of the data.

13 **a** $E(X) = 3$.
$\text{Var}(X) = (5 + 1)(5 - 1)/12 = 2$.
b 7
c 18

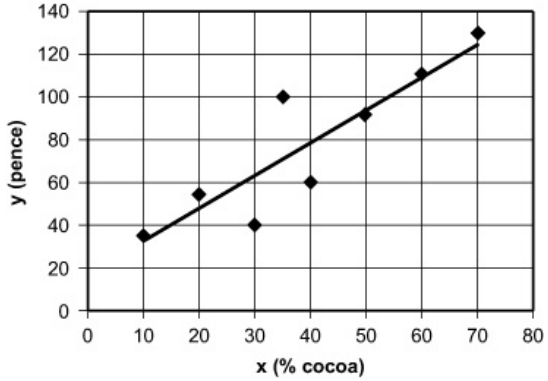
14 **a**



b mean 1.700 m, standard deviation 0.095 m
c 0.337

15 **a** -0.816
b Houses are cheaper further away from the town centre.
c -0.816

16 **a & d**



b $S_{xy} = 28\,750 - (315 \times 620)/8 = 4337.5$,
$S_{xx} = 2822$
c $a = 17.0$, $b = 1.54$
e **i** Brand D is overpriced, since it is a long way above the line.
ii 69 *or* 70 pence

17 **a** 0.1056
b 27.2

18 **a** $p + q = 0.45$, $3p + 7q = 1.95$.
b $p = 0.3$, $q = 0.15$.
c 0.35
d 7.15
e 1
f 114.4

19 **a** 0.5
b 54 kg
c It represents an outlier or extreme value. It could be drums or a double base.
d No skew.
e 13.3 kg *or* 13.4 kg

Exam Style Paper Solutions

Mark scheme

M marks are awarded for knowing the method and attempting to use it.
A marks are given for appropriately accurate correct answers.
A marks are not awarded without the method marks.
B marks are given for correct answers.

1.

3	*4*	*4*	*4*	*5*	*5*	*6*
2	*3*	*3*	*3*	*4*	*4*	*5*
2	*3*	*3*	*3*	*4*	*4*	*5*
1	2	2	2	*3*	*3*	*4*
1	2	2	2	*3*	*3*	*4*
1	2	2	2	*3*	*3*	*4*
Second First	**1**	**1**	**1**	**2**	**2**	**3**

The easiest solution involves drawing a diagram to represent the sample space. Each square is the sum of the scores on the die. The first method mark is for attempting the diagram and the second is an accuracy mark for all the values correct.

Each of the values that are 'at least 3' are underlined; 3, 4, 5 & 6.

M1A1A1

P(Sum at least 3)$=\frac{27}{36}=\frac{3}{4}$

There are 27 values underlined and 36 values in the sample space. Then cancel the fraction.

M1A1

ALTERNATIVE SOLUTION

This is a slightly quicker solution.

Let D_1 = the number on the first die
and D_2 = the number on the second die

$$\begin{aligned} P(D_1 + D_2 \geq 3) &= 1 - P(D_1 + D_2 = 2) \\ &= 1 - P(D_1 = 1 \text{ and } D_2 = 1) \\ &= 1 - P(D_1 = 1) \times P(D_2 = 1) \\ &= 1 - \tfrac{1}{2} \times \tfrac{1}{2} \\ &= \frac{3}{4} \end{aligned}$$

P(D=1) = 0.5 and D_1 and D_2 are independent so the probabilities are multiplied together.

2. (a) $P(X < 450) = P\left(Z < \frac{450-460}{10}\right) = P(Z<-1.0)$ M1A1

$= 1 - 0.8413 = 0.1587$ A1

Standardise by subtracting the mean and dividing by the standard deviation gets the first method mark and the z value of -1.0 gets the accuracy mark.

(b) Expected number of jars = 30×0.1587 M1

$= 4.761$ or 4.76 or 4.8 A1

(c) $P(X < 450) = 0.01$ M1

$\frac{450-\mu}{10} = -2.3263$ M1A1

$\mu = 473.263 = 473$ to 3 sf A1

Forming the correct equation with the new mean as an unknown gets the method and accuracy mark, the B mark is awarded for getting -2.3263 from the tables

---

3. (a) $0.5 + b + 2a = 1$ B1

$0.3 + 2b + 6a = 1.6$ M1A1

Solving
$a = 0.15, b = 0.2$ M1A1

Remember that adding all the probabilities together equals 1.

The second equation is formulated from the value of the expectation. Multiply the values of X by the associated probabilities and equate to 1.6.

(b) $E(5-2X) = 5 - 2E(X)$ M1

$= 5 - 2 \times 1.6 = 1.8$ A1

(c) $Var(X) = 1^2 \times 0.3 + 2^2 \times 0.2 + 3^2 \times 0.3 - 1.6^2$ M1A1

$= 1.24$ A1

For the variance you square each value of x and multiply by the probability. Remember to subtract the square of the expectation.

---

4. (a) $\bar{x} = \frac{302}{16} = 18.875$

standard deviation is $\sqrt{\frac{5722}{16} - 18.875^2} = \sqrt{1.359375}$

$= 1.16592\ldots$

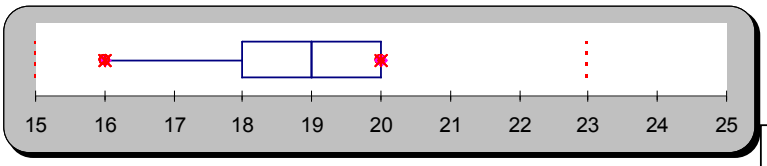
(b) mean % attendance is $\frac{18.875}{20} \times 100 = 94.375$

(c) Mode is 17

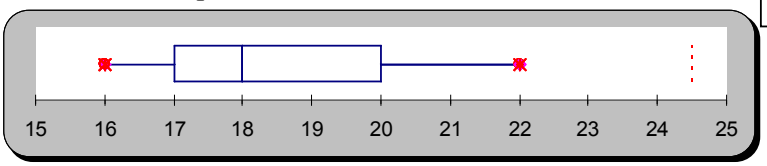
Median is 18

IQR is $20 - 17 = 3$

(d) First Group:



Second Group:



(e) First mean % > Second mean %
First IQR < Second IQR
First sd < Second sd
First range < Second range
First negative skew, given by whiskers, symmetric by box
Second positive skew.

Set out your working clearly so you will still be given the method mark if you make a calculator error.

Put the box plots side by side so you can compare easily.

There are 3 marks for this part, so 3 different correct comments are required. Try to comment about location, spread and shape.

M1A1

M1

A1

B1

B1

B1

B1

M1A1
A1

B1B1B1

5. (a) Discrete uniform distribution

B1

(b) $\frac{(n+1)}{2}=10$

$n=19$

M1

A1

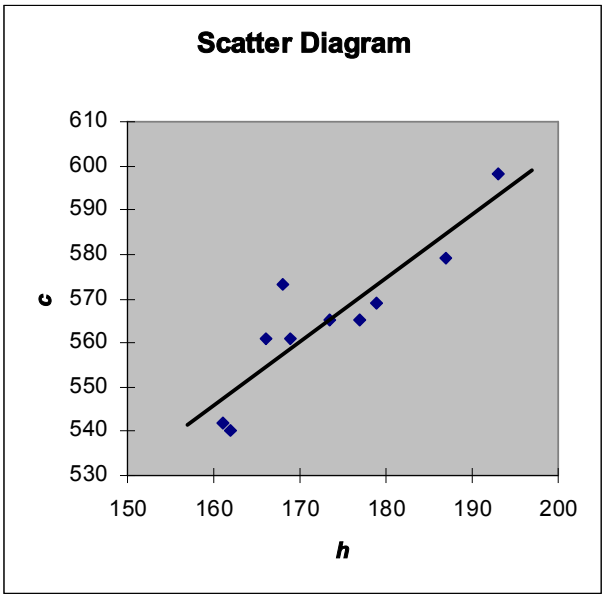
Learning the details of the uniform distribution and formulae for mean and variance make this question easier.

(c) $\frac{(n+1)(n-1)}{2}=180$

M1A1

---

6. (a) & (e)



Be careful when plotting the points.
Make sure the regression line passes through this point. $(\bar{h},\bar{c})$

B1B1 (2) for points, B1B1 (2) for line.

(b) $S_{hh}=272094-\frac{1562^2}{9}=1000.\dot{2}$

B1

$S_{cc}=2878966-\frac{5088^2}{9}=2550$

B1

$S_{hc}=884484-\frac{1562\times5088}{9}=1433.\dot{3}$

B1

Make sure you work accurately as all these marks are for the answers.

(c) $r=\frac{S_{hc}}{\sqrt{S_{hh}S_{cc}}}=\frac{1433.\dot{3}}{\sqrt{1000.\dot{2}\times2550}}=0.897488$

Don't forget the square root.

M1A1A1

(d) $b = \dfrac{1433.\dot{3}}{1000.\dot{2}} = 1.433015$

$a = \dfrac{5088}{9} - b \times \dfrac{1562}{9} = 316.6256$

$c = 1.43h + 317$

B1

Set out your working carefully and remember to minus the b in the second equation.

M1A1

A1

(e) See Graph

(a) For every 1cm increase in height, the confidence measure increases by 1.43.

This must be in context i.e. it relates to 'height' and 'confidence measure'.

(g) $h = 172$

$c = 1.43 \times 172 + 317 = 563$ to 3 sf

Substituting h = 172 into your equation gets the method mark.

---

7. (a) P(Scores 15 points)
= P(hit,hit,hit)=$0.4 \times 0.4 \times 0.4 = 0.064$

There is only one way of scoring 15 points.

M1A1

(b)

x	0	5	10	15
P(X=x)	0.6	0.4×0.6	$0.4^2 \times 0.6$	
	0.6	0.24	0.096	0.064

Set out the distribution in a table

B1

(c) P(Jean scores more in round two than round one)
=P(X = 0 then X = 5, 10 or 15)
+ P(X = 5 then X = 10 or 15)
+ P(X = 10 then X = 15)

=0.6 × (0.24 + 0.096 + 0.064)
+0.24 × (0.096 + 0.064)
+0.096 × 0.064

=0.284544
= 0.285 (3 sf)

There is only 1 way of scoring each value as the round ends if Jean misses.

M1A1

Consider the possible score for the first round in turn and the corresponding scores on the second round.

A1

A1

A1

A1
