Review exercise

Paper 1 style questions

EXAM-STYLE QUESTIONS

1 It is stated that the content of a can of drink is 350 ml. The content of thousands of cans is tested and found to be normally distributed with a mean of 354 ml and a standard deviation of 2.5 ml.

a Sketch a normal distribution diagram to illustrate this information.

b Find the probability that a can contains less than 350 ml.

100 cans are chosen at random.

c Find the expected number of cans that contain less than 350 ml.

2 6000 people were asked how far they lived from their work. The distances were normally distributed with a mean of 4.5 km and a standard deviation of 1.5 km.

a Find the percentage of people who live between 2 km and 4 km from their work.

b Find the expected number of people who live less than 1 km from their work.

3 The weights of bags of tomatoes are normally distributed with a mean of 1.03 kg and a standard deviation of 0.02 kg.

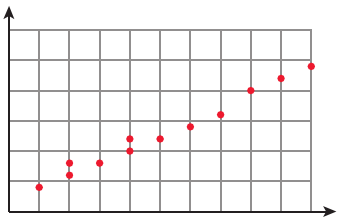
a Find the percentage of bags that weigh more than 1 kg.

It is known that 15% of the bags weigh less than p kg.

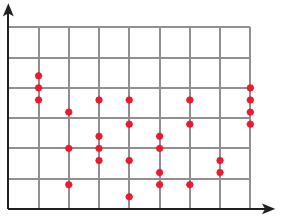
b Find the value of p.

4 For each diagram, state the type of correlation.

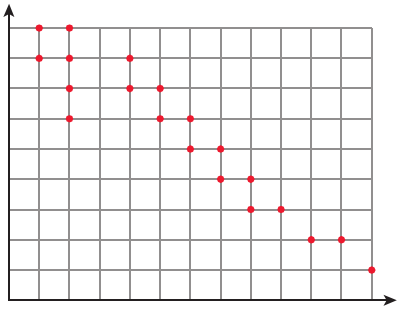
a



b



c



5 Plot these points on a diagram.

x	6	8	10	12	14	16
y	20	21	24	27	28	30

a State the nature of the correlation.

b Find the mean of the x-values and the mean of the y-values. Plot this mean point on your diagram.

c Draw the line of best fit by eye.

d Find the expected value for y when $x = 9$.

6 The heights and arm lengths of 10 people are shown in the table.

Height (cm)	145	152	155	158	160	166	172	179	183	185
Arm length (cm)	38	42	45	53	50	59	61	64	70	69

a Find the correlation coefficient, r, and comment on your result.

b Write down the equation of the regression line.

c Use your equation to estimate the arm length of a person of height 170 cm.

7 The time taken to eat three doughnuts and the person's age is recorded in the table.

Age (years)	8	12	15	18	21	30	33	35	44	52	63	78
Time (seconds)	23	21	17	14	15	18	20	21	23	25	27	35

a Find the correlation coefficient, r, and comment on your result.

b Write down the equation of the regression line.

c Use your equation to estimate the time taken by a 40-year-old to eat three doughnuts.

8 100 people are asked to identify their favorite flavor of ice cream. The results are given in the contingency table, classified by age (x).

	$x < 25$	$25 \leq x < 45$	$x \geq 45$	**Totals**
Vanilla	14	13	10	37
Strawberry	11	9	8	28
Chocolate	13	10	12	35
Totals	38	32	30	100

Perform a chi-squared test, at the 5% significance level, to determine whether flavor of ice cream is independent of age. State clearly the null and alternative hypotheses, the expected values and the number of degrees of freedom.

9 60 students go ten-pin bowling. They each have one throw with their right hand and one throw with their left. The number of pins knocked down each time is noted. The results are collated in the table.

	0–3	4–7	8–10	Totals
Right hand	8	28	24	60
Left hand	12	30	18	60
Totals	20	58	42	120

A χ^2 test is performed at the 10% significance level.

a State the null hypothesis.

b Write down the number of degrees of freedom.

c Show that the expected number of students who knock down 0–3 pins with their right hand is 10.

The p-value is 0.422.

d Write down the conclusion reached at the 10% significance level.
Give a clear reason for your answer.

10 Erland performs a chi-squared test to see if there is any association between the preparation time for a test (short time, medium time, long time) and the outcome (pass, does not pass). Erland performs this test at the 5% significance level.

a Write down the null hypothesis.

b Write down the number of degrees of freedom.

The p-value for this test is 0.069.

c What conclusion can Erland make?
Justify your answer.

Paper 2 style questions

EXAM-STYLE QUESTIONS

1 The heights of Dutch men are normally distributed with a mean of 181 cm and a standard deviation of 9 cm.

a Sketch a normal distribution diagram to illustrate this information.

b Find the probability that a man chosen at random has a height less than 175 cm.

c Find the probability that a man chosen at random has a height between 172 cm and 192 cm.

Sixty men are measured.

d Find the expected number of men with a height greater than 195 cm.

It is known that 5% of the men have a height less than k cm.

e Find the value of k.

2 The weights of bags of sweets are normally distributed with a mean of 253 g and a standard deviation of 3 g.

a Sketch a diagram to illustrate this information clearly.

b Find the percentage of bags expected to weigh less than 250 g.

Three hunderd bags are weighed.

c Find the expected number of bags weighing more than 255 g.

3 The heights and weights of 10 students selected at random are shown in the table.

Height (x cm)	158	167	178	160	152	160	173	181	185	155
Weight (y kg)	50	75	80	46	61	69	64	86	74	68

a Plot this information on a scatter graph. Use a scale of 1 cm to represent 25 cm on the x-axis and 1 cm to represent 10 kg on the y-axis.

b Calculate the mean height.

c Calculate the mean weight.

d **i** Find the equation of the regression line.

ii Draw the regression line on your graph.

e Use your line to estimate the weight of a student of height 170 cm.

4 An employment agency has a new computer software package. The agency investigates the number of hours it takes people of different ages to reach a satisfactory level using this package. Fifteen people are tested and the results are given in the table.

Age (x)	33	41	22	46	25	18	16	23	26	55	37	34	25	48	17
Time (y hours)	8	10	7	16	8	9	7	10	12	15	11	14	10	16	7

a Find the product-moment correlation coefficient, r, for these data.

b What does the value of the correlation coefficient suggest about the relationship between the two variables?

c Write down the equation of the regression line for y on x in the form $y = mx + c$.

d Use your equation for the regression line to predict the time that it would take a 35-year-old person to reach a satisfactory level. Give your answer correct to the nearest hour.

EXAM-STYLE QUESTIONS

5 Ten students were asked for their average grade at the end of their last year of high school and their average grade at the end of their first year at university. The results were put into a table as follows.

Student	High school grade, x	University grade, y
1	92	3.8
2	76	2.9
3	83	3.4
4	71	1.8
5	93	3.9
6	84	3.2
7	96	3.5
8	77	2.9
9	91	3.7
10	86	3.8

a Find the correlation coefficient, r, giving your answer to one decimal place.

b Describe the correlation between the high school grades and the university grades.

c Find the equation of the regression line for y on x in the form $y = mx + c$.

6 Several bars of chocolate were purchased and the following table shows the weight and the cost of each bar.

	Yum	Choc	Marl	Twil	Chuns	Lyte	BigM	Bit
Weight (x grams)	58	75	70	68	85	52	94	43
Cost (y euros)	1.18	1.45	1.32	1.05	1.70	0.90	1.53	0.95

a Find the correlation coefficient, r, giving your answer correct to two decimal places.

b Describe the correlation between the weight of a chocolate bar and its cost.

c Calculate the equation of the regression line for y on x.

d Use your equation to estimate the cost of a chocolate bar weighing 80 g.

7 The heights and dress sizes of 10 female students selected at random are shown in the table.

Height (x cm)	175	160	180	155	178	159	166	185	189	173
Dress size (y)	12	14	14	8	12	10	14	16	16	14

a Write down the equation of the regression line for dress size (y)
on height (x), giving your answer in the form $y = ax + b$.

b Use your equation to estimate the dress size of a student of height 170 cm.

c Write down the correlation coefficient.

d Describe the correlation between height and dress size.

EXAM-STYLE QUESTIONS

8 Members of a certain club are required to register for one of three games: badminton, table tennis or darts.
The number of club members of each gender choosing each game in a particular year is shown in the table.

	Badminton	Table tennis	Darts
Male	37	16	28
Female	32	10	19

Use a chi-squared test, at the 5% significance level, to test whether choice of game is independent of gender. State clearly the null and alternative hypotheses, the expected values and the number of degrees of freedom.

9 For his Mathematical Studies Project a student gave his classmates a questionnaire to find out which extra-curricular activity was the most popular. The results are given in the table below, classified by gender.

	Reading	Surfing	Skating	
Female	22	16	22	(60)
Male	14	18	8	(40)
	(36)	(34)	(30)	

The table below shows the expected values.

	Reading	Surfing	Skating
Female	p	20.4	18
Male	q	r	12

a Calculate the values of p, q and r.

The chi-squared test, at the 10% level of significance, is used to determine whether the extra-curricular activity is independent of gender.

b i State a suitable null hypothesis.

ii Show that the number of degrees of freedom is 2.

The critical value is 4.605.

c Write down the chi-squared statistic.

d Do you accept the null hypothesis? Explain your answer.

EXAM-STYLE QUESTIONS

10 A company conducted a survey to determine whether position in upper management was independent of gender. The results of this survey are tabulated below.

	Managers	Junior executives	Senior executives	Totals
Male	135	90	75	300
Female	45	130	25	200
Totals	180	220	100	500

The table below shows the expected number of males and females at each level, if they were represented proportionally to the total number of males and females employed.

	Managers	Junior executives	Senior executives	Totals
Male	*a*	*c*	60	300
Female	*b*	*d*	40	200
Totals	180	220	100	500

a **i** Show that the expected number of male managers (a) is 108.

ii Hence, write down the values of b, c and d.

b Write suitable null and alternative hypotheses for these data.

c **i** Find the chi-squared value.

ii Write down the number of degrees of freedom.

iii Given that the critical value is 5.991, what conclusions can be drawn regarding gender and position in upper management?

11 In the small town of Schiedam, population 8000, an election was held. The results were as follows.

	Urban voters	Rural voters
Candidate A	1950	1730
Candidate B	1830	1360
Candidate C	500	630

In **a**–**d** below, use a chi-squared test, at the 1% significance level, to decide whether the choice of candidate depends on where the voter lives.

H_0: The choice of candidate is independent of where the voter lives.

a Write down the alternative hypothesis.

b Show that the expected number of rural voters for candidate A is 1711.

c **i** Calculate the chi-squared value.

ii Write down the number of degrees of freedom.

The critical value is 9.21.

d **i** State your conclusion.

ii Explain why you reached your conclusion.

EXAM-STYLE QUESTION

12 This table of observed results gives the number of candidates taking a Mathematics examination classified by gender and grade obtained.

	6 or 7	4 or 5	1, 2 or 3	Totals
Males	34	50	6	90
Females	40	60	10	110
Totals	74	110	16	200

The question posed is whether gender and grade obtained are independent.

a Show that the expected number of males achieving a grade of 4 or 5 is 49.5.

A chi-squared test is set up at the 5% significance level.

b **i** State the null hypothesis.

ii State the number of degrees of freedom.

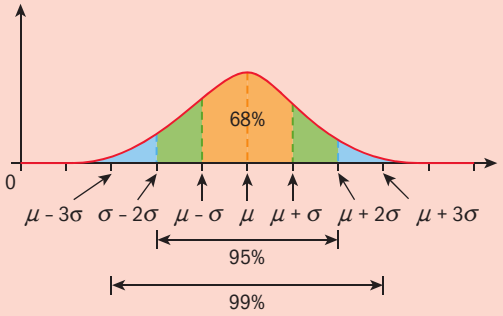
iii Write down the chi-squared value.

The critical value is 5.991.

c What can you say about gender and grade obtained?

CHAPTER 5 SUMMARY
The normal distribution

- The **normal distribution** is the most important continuous distribution in statistics. It has these properties:
  - It is a bell-shaped curve.
  - It is symmetrical about the mean, μ. (The mean, the mode and the median all have the same value.)
  - The x-axis is an asymptote to the curve.
  - The total area under the curve is 1 (or 100%).
  - 50% of the area is to the left of the mean and 50% to the right.
  - Approximately 68% of the area is within 1 standard deviation, σ, of the mean.
  - Approximately 95% of the area is within 2 standard deviations of the mean.
  - Approximately 99% of the area is within 3 standard deviations of the mean.

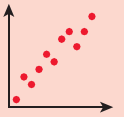


- The **expected value** is found by multiplying the number in the sample by the probability.

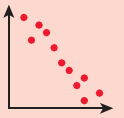
Continued on next page

Correlation

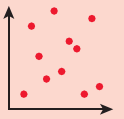
- In a **positive** correlation the dependent variable increases as the independent variable increases.



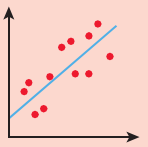
- In a **negative** correlation the dependent variable decreases as the independent variable increases.



- When the points are scattered randomly across the diagram there is **no** correlation.



- Correlations can also be described as strong, moderate or weak.
- To draw the **line of best** fit by eye:
  - Find the mean of each set of data and plot this point on your scatter diagram.
  - Draw a line that passes through the mean point and is close to all the other points – with approximately an equal number of points above and below the line.



- **Pearson's product-moment correlation coefficient**, r, can take all values between -1 and $+1$ inclusive.
  - When $r = -1$, there is a **perfect negative** correlation between the data sets.
  - When $r = 0$, there is **no** correlation.
  - When $r = +1$, there is a **perfect positive** correlation between the data sets.
  - A **perfect correlation** is one where **all** the plotted points lie on a straight line.

The regression line

- The **regression line for *y* on *x*** is a more accurate version of a line of best fit, compared to best fit by eye.
- If there is a strong or moderate correlation, you can use the regression line for y on x to predict values of y for values of x within the range of the data.

The chi-squared test

- To calculate the χ^2 value use the formula $\chi^2_{\text{calc}} = \sum \frac{(f_o - f_e)^2}{f_e}$, where f_o are the observed frequencies and f_e are the expected frequencies.
- To find the degrees of freedom for the chi-squared test for independence, use this formula based on the contingency table:
  Degrees of freedom = (number of rows − 1)(number of columns − 1)

Continued on next page

- If χ^2_{calc} is **less than** critical value, **do not reject** the null hypothesis.
  If χ^2_{calc} is **more than** critical value, **reject** the null hypothesis.
- If the *p*-value is **less than** significance level, **reject** the null hypothesis.
  If the *p*-value is **more than** significance level, **do not reject** the null hypothesis.
- To perform a χ^2 test:
  1. Write the null (H_0) and alternative (H_1) hypotheses.
  2. Calculate χ^2_{calc}: **a** using your GDC (examinations), or **b** using the χ^2_{calc} formula (project work).
  3. Determine: **a** the *p*-value using your GDC, or **b** the critical value (given in examinations).
  4. Compare: **a** the *p*-value against the significance level, or **b** χ^2_{calc} against the critical value.

7 a $x = -2$ b (0, 0), (–4, 0)
c (–2, –4)

8 a $x = -\frac{1}{2}$ b (0, 0), (–1, 0)
c $\left(-\frac{1}{2}, -\frac{1}{4}\right)$

9 a $x = 1$ b (–1, 0), (3, 0)
c (1, –4)

10 a $x = 1$ b (5, 0), (–3, 0)
c (1, –16)

11 a $x = 4$ b (2, 0), (6, 0)
c (4, –4)

12 a $x = 1$ b (4, 0), (–2, 0)
c (1, –9)

Investigation – the general quadratic curve $y = ax^2 + bc + c$

Part A

1 The curve intersects the x-axis at (1, 0) and (3, 0).

The axis of symmetry is at $x = 2$
The co-ordinate of the vertex is (2, –1)

2 In the case $a = 1$:

The curve intersects the x-axis at $\left(\frac{-b - \sqrt{b^2 - 4ac}}{2a}, 0\right)$ and $\left(\frac{-b + \sqrt{b^2 - 4ac}}{2a}, 0\right)$.

The axis of symmetry is at $x = \frac{-b}{2}$

The co-ordinate of the vertex is $\left(\frac{-b}{2}, c - \frac{-b}{4}\right)$

Part B

1 The curve does not intercept the x-axis

The axis of symmetry is at $x = 1$
The co-ordinate of the vertex is (1, 1)

Exercise 4L

1 a $x = 1$ b no points
c (1, 2)

2 a $x = -2$ b (1, 0), (– 5, 0)
c (–2, –9)

3 a $x = -3$
b (– 0.764, 0), (–5.24, 0)
c (–3, –5)

4 a $x = 1$ b (0.423, 0), (1.58, 0)
c (1, –1)

5 a $x = 2$
b (– 0.121, 0), (4.12, 0)
c (2, –9)

6 a $x = -\frac{3}{2}$
b (0.898, 0), (– 3.90, 0)
c $\left(-\frac{3}{2}, -\frac{23}{2}\right)$

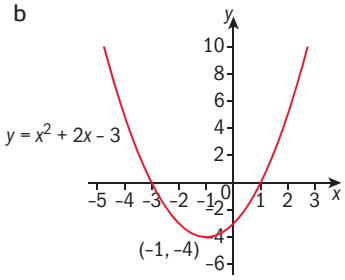
7 a $x = 1$
b no points
c $\left(1, \frac{3}{2}\right)$

8 a $x = -3$
b (1.12, 0), (– 7.12, 0)
c $\left(-3, -\frac{17}{2}\right)$

Exercise 4M

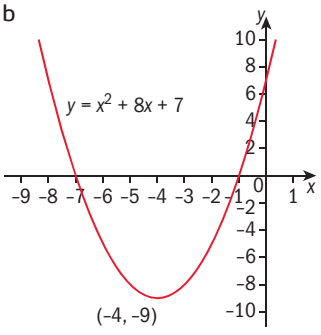
1 a i (0, –3) ii $x = -1$
iii (– 1, – 4) iv (– 3, 0), (1, 0)
v $y \geq -4$

b



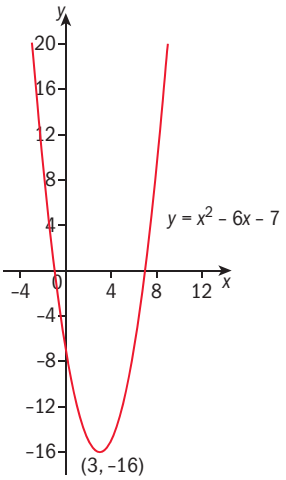
2 a i (0, 7) ii $x = -4$
iii (–4, –9) iv (–7, 0), (–1, 0)
v $y \geq -9$

b



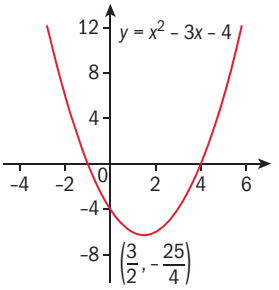
3 a i (0, – 7) ii $x = 3$
iii (3, –16)
iv (7, 0), (– 1, 0)
v $y \geq -16$

b



4 a i (0, –4) ii $x = \frac{3}{2}$
iii $\left(\frac{3}{2}, -\frac{25}{4}\right)$
iv (4, 0), (– 1, 0)
v $y \geq -\frac{25}{4}$

b



5 a i (0, –10) ii $x = \frac{3}{2}$
iii $\left(\frac{3}{2}, -\frac{49}{4}\right)$
iv (5, 0), (– 2, 0)
v $y \geq -\frac{49}{4}$

b