SL Paper 1

A researcher consulted 500 men and women to see if the colour of the car they drove was independent of gender. A χ^2 test for independence was carried out.

a. Write down the null hypothesis. [1]

b. The colours of the cars were red, green, blue, black and silver. [2]

Find the number of degrees of freedom for this test.

c. At the 5 % significance level the χ^2_{calc} was found to be 8.73. [1]

Write down the critical value, χ^2_{crit}, for this test.

Markscheme

a. Colour of car is independent of gender. (Colour of car and gender are independent) ***(A1)*** ***(C1)***

Note: Accept "not associated". Do not accept "not related", "not correlated" or "not linked".

[1 mark]

b. (2 – 1)(5 – 1) = 4 ***(M1)(A1)*** ***(C2)***

[2 marks]

c. $\chi^2_{crit} = 9.488$ ***(A1)*****(ft)** ***(C1)***

Notes: Accept 9.49. Follow through from part (b).

[1 mark]

Examiners report

a. This question was well answered by the majority of the candidates, many scoring the maximum number of marks. It was disappointing to see the number of candidates who left this question blank.

b. This question was well answered by the majority of the candidates, many scoring the maximum number of marks. It was disappointing to see the number of candidates who left this question blank.

c. This question was well answered by the majority of the candidates, many scoring the maximum number of marks. It was disappointing to see the number of candidates who left this question blank.

180 spectators at a swimming championship were asked which, of four swimming styles, was the one they preferred to watch.

The results of their responses are shown in the table.

Swimming style	Male	Female
Freestyle	20	15
Butterfly	20	30
Backstroke	10	35
Breaststroke	10	40

A χ^2 test was conducted at the 5% significance level.

a. Write down the null hypothesis for this test. [1]

b. Write down the number of degrees of freedom. [1]

c. Write down the value of χ^2_{calc}. [2]

d. The critical value, at the 5% significance level, is 7.815. [2]

State, giving a reason, the conclusion to the test.

Markscheme

a. The (preferred) swimming style is independent of gender ***(A1)*** ***(C1)***

Notes: Accept “not associated”. Do not accept “not related”, “not correlated” or “not influenced”.

[1 mark]

b. 3 ***(A1)*** ***(C1)***

[1 mark]

c. $\chi^2_{calc} = 16.4\ (16.4285\ldots)$ ***(A2)*** ***(C2)***

[2 marks]

d. Do not accept the Null Hypothesis (Reject the Null Hypothesis).

$\chi^2_{calc} > \chi^2_{crit}$ **OR** $16.4 > 7.815$ ***(R1)(A1)*****(ft)**

OR

Do not accept the Null Hypothesis (Reject the Null Hypothesis).

p-value of $9.26148\ldots \times 10^{-4} < 0.05$ ***(R1)(A1)*****(ft)** ***(C2)***

Notes: Follow through from their answer to part (c).

Accept “(preferred) swimming style is not independent (dependent) of gender” as the conclusion.

Do not award ***(R0)(A1)***.

If using the p-value the value must be seen.

[2 marks]

Examiners report

a. This question was well answered by the majority of the candidates, many scoring the mark.

b. This question was well answered by the majority of the candidates, many scoring the mark.

c. This question was well answered by the majority of the candidates, many scoring both marks.

d. This question was well answered by the majority of the candidates, many scoring both marks.

The heights of apple trees in an orchard are normally distributed with a mean of $3.42\ \text{m}$ and a standard deviation of $0.21\ \text{m}$.

a. Write down the probability that a randomly chosen tree has a height greater than $3.42\ \text{m}$. [1]

b. Write down the probability that a randomly chosen tree will be within 2 standard deviations of the mean of $3.42\ \text{m}$. [1]

c. Use your graphic display calculator to calculate the probability that a randomly chosen tree will have a height greater than $3.35\ \text{m}$. [2]

d. The probability that a particular tree is less than x metres high is 0.65. Find the value of x. [2]

Markscheme

a. $0.5\ \left(50\%,\ \frac{50}{100},\ \frac{1}{2}\right)$ ***(A1)*** ***(C1)***

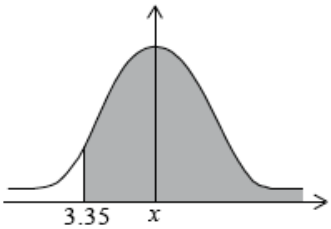
[1 mark]

b. $0.954 (0.954499\ldots, 95.4\%, 95.4499\ldots\%)$ ***(A1)*** ***(C1)***

Note: Accept 95% or 0.95.

[1 mark]

c.

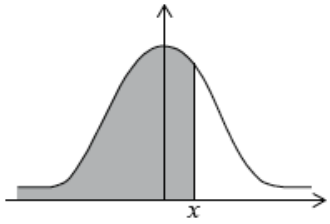


(M1)

Note: Accept alternative methods.

$0.631 (0.630558\ldots, 63.1\%, 63.0558\ldots\%)$ ***(A1)*** ***(C2)***

[2 marks]

d. 

(M1)

Note: Accept alternative methods.

$3.50\ (3.50091...)$ ***(A1)*** ***(C2)***

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

Members of a certain club are required to register for one of three sports, badminton, volleyball or table tennis. The number of club members of each gender choosing each sport in a particular year is shown in the table below.

A χ^2 (Chi-squared) test at the 5% significance level is used to determine whether the choice of sport is independent of gender.

	Badminton	Volleyball	Table tennis
Male	40	20	10
Female	20	15	15

a. Find the expected number of female volleyball players under this hypothesis. [2]

b. Write down the p-value for the test. [2]

c. State, with a reason, the conclusion of the test. [2]

Markscheme

a. $\frac{50}{120} \times \frac{35}{120} \times 120$ **OR** $\left(\frac{50\times35}{120}\right)$ ***(M1)***

$= 14.6\ (14.5833\ldots)$ ***(A1)*** ***(C2)***

b. 0.0746 ***(A2)*** ***(C2)***

c. Since $p\text{-value} > 5\%$, the choice of the sport is independent of gender. ***(R1)(A1)*****(ft)** ***(C2)***

Note: The ***(R1)*** is awarded for the explicit comparison, the ***(A1)*(ft)** is awarded for a consistent conclusion with their answer in part (c).
It is therefore possible that ***(R1)(A0)*** may be awarded, but ***(R0)(A1)*** can never be awarded.

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

Tony wants to carry out a χ^2 test to determine whether or not a person's choice of one of the three professions; engineering, medicine or law is influenced by the person's sex (gender).

a. State the null hypothesis, H_0, for this test. [1]

b. Write down the number of degrees of freedom. [1]

c. Of the 400 people Tony interviewed, 220 were male and 180 were female. 80 of the people had chosen engineering as a profession. [2]

Calculate the expected number of female engineers.

d. Tony used a 5 % level of significance for his test and obtained a *p*-value of 0.0634 correct to 3 significant figures. [2]

State Tony's conclusion to the test. Give a reason for this conclusion.

Markscheme

a. Chosen profession is independent of gender. ***(A1)***

OR

There is no association between gender and chosen profession. ***(A1)*** ***(C1)***

Note: Do not accept "not related", "not correlated" or "not influenced".

[1 mark]

b. 2 ***(A1)*** ***(C1)***

[1 mark]

c. $\frac{180\times80}{400}$ ***(M1)***

OR

$\frac{180}{400}\times\frac{80}{400}\times400$ ***(M1)***

$=36$ ***(A1)*** ***(C2)***

[2 marks]

d. *p*-value > 0.05 ***(R1)***

Accept H_0 ***(A1)*** ***(C2)***

Note: Do not award ***(R0)(A1)***.

[2 marks]

Examiners report

a. The first two parts of this question were very well answered but a number of students found calculating the required expected value in part c) difficult. Very few knew how to use the given *p*-value in order to decide whether to reject or retain the null hypothesis. There were some candidates who did not attempt this question at all which might be indicating that this topic had not been discussed in some schools.

b. The first two parts of this question were very well answered but a number of students found calculating the required expected value in part c) difficult. Very few knew how to use the given *p*-value in order to decide whether to reject or retain the null hypothesis. There were some candidates who did not attempt this question at all which might be indicating that this topic had not been discussed in some schools.

c. The first two parts of this question were very well answered but a number of students found calculating the required expected value in part c) difficult. Very few knew how to use the given *p*-value in order to decide whether to reject or retain the null hypothesis. There were some candidates who did not attempt this question at all which might be indicating that this topic had not been discussed in some schools.

d. The first two parts of this question were very well answered but a number of students found calculating the required expected value in part c) difficult. Very few knew how to use the given *p*-value in order to decide whether to reject or retain the null hypothesis. There were some candidates who did not attempt this question at all which might be indicating that this topic had not been discussed in some schools.

The producer of a TV dancing show asked a group of 150 viewers their age and the type of Latin dance they preferred. The types of Latin dances in the show were Argentine tango, Samba, Rumba and Cha-cha-cha. The data obtained were organized in the following table.

	Dance			
	Argentine tango	**Samba**	**Rumba**	**Cha-cha-cha**
20 years old and younger	35	23	12	10
Older than 20 years old	20	17	18	15

A χ^2 test was carried out, at the 5% significance level.

a. Write down the null hypothesis for this test. [1]

b. Write down the observed number of viewers who preferred Rumba **and** were older than 20 years old. [1]

c. Use your graphic display calculator to find the p-value for this test. [2]

d. The producer claims that the type of Latin dance a viewer preferred is independent of their age. [2]

Decide whether this claim is justified. Give a reason for your decision.

Markscheme

a. H_0 the type of Latin dance the viewer prefers is independent of their age ***(A1)*** ***(C1)***

Notes: Accept “not dependent” or “not associated”. Do not accept “not correlated” or “not related” or “not connected”.

b. 18 ***(A1)*** ***(C1)***

c. $p = 0.0876 \quad (0.0875813\ldots)$ ***(A2)*** ***(C2)***

Notes: Award ***(A2)*** for 0.088.

Award ***(A1)(A0)*** for an answer of 0.0875.

d. $0.05 <$ their p-value ***(R1)***

the producer’s claim is justified ***(A1)*(ft)*** ***(C2)***

Notes: Do not award ***(R0)(A1)*(ft)***. Follow through from their answer to (c). If there is no answer in part (c), award ***(R1)(A0)*** for stating the relationship between the independence **and** the p-value compared to 0.05.

If ***(R1)*** is awarded, award ***(A1)*(ft)*** for the answer ‘yes’ or the answer ‘no’ if it is consistent with their reasoning.

Similarly, allow ‘accept H_0’ or ‘reject H_0’ if consistent with their reasoning.

Award ***(R0)*** for comparing p with the critical value.

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

A market researcher consulted males and females to determine whether the type of coffee they drink is associated with gender. The types of coffee are Cappuccino, Latte, Americano, Macchiato and Espresso. A χ^2 test was conducted, at the 5 % significance level and the χ^2 value was found to be 8.73.

a.i. Write down the null hypothesis. [1]

a.ii. Write down the alternative hypothesis. [1]

b. Write down the number of degrees of freedom for this test. [1]

c. Write down the critical value for this test. [1]

d. State whether the type of coffee drunk is independent of gender. Give a reason for your answer. [2]

Markscheme

a.i. Type of coffee drunk is **independent** of gender. ***(A1)***

Note: Accept is “not associated”. Do not accept “not related”, “not correlated” or “not influenced”.

[1 mark]

a.ii. Type of coffee drunk is **not independent** of gender. ***(A1)*****(ft)** ***(C2)***

Note: If hypotheses are reversed award ***(A0)(A1)*****(ft)**.

[1 mark]

b. 4 ***(A1)*** ***(C1)***

[1 mark]

c. $\chi^2(\text{crit}) = 9.488$ ***(A1)*****(ft)** ***(C1)***

Note: Accept 9.49.

[1 mark]

d. $\chi^2(\text{calc}) < \chi^2(\text{crit}),\ 8.73 < 9.488$ ***(R1)***

Accept the null hypothesis (Accept type of coffee is independent of gender). ***(A1)*****(ft)** ***(C2)***

Notes: Follow through from their answer to part (c).

Do not award ***(R0)(A1)***.

[2 mark]

Examiners report

a.i. Many candidates gained full marks on this question. However, a number of candidates did not answer at all or stopped after either correctly or incorrectly defining H_0 and/or H_1. Many incorrect versions of ‘independent’ were seen and candidates should be advised that the terms not related, not correlated and not influenced will not be awarded marks. There were an encouraging number of full marks gained on this question.

a.ii. Many candidates gained full marks on this question. However, a number of candidates did notanswer at all or stopped after either correctly or incorrectly defining H_0 and/or H_1. Many incorrect versions of ‘independent’ were seen and candidates should be advised that the termsnot related, not correlated and not influenced will not be awarded marks. There were anencouraging number of full marks gained on this question.

b. Many candidates gained full marks on this question. However, a number of candidates did not answer at all or stopped after either correctly or incorrectly defining H_0 and/or H_1. Many incorrect versions of 'independent' were seen and candidates should be advised that the terms not related, not correlated and not influenced will not be awarded marks. There were an encouraging number of full marks gained on this question.

c. Many candidates gained full marks on this question. However, a number of candidates did not answer at all or stopped after either correctly or incorrectly defining H_0 and/or H_1. Many incorrect versions of 'independent' were seen and candidates should be advised that the terms not related, not correlated and not influenced will not be awarded marks. There were an encouraging number of full marks gained on this question.

d. Many candidates gained full marks on this question. However, a number of candidates did not answer at all or stopped after either correctly or incorrectly defining H_0 and/or H_1. Many incorrect versions of 'independent' were seen and candidates should be advised that the terms not related, not correlated and not influenced will not be awarded marks. There were an encouraging number of full marks gained on this question.

A study was carried out to determine whether the country chosen by students for their university studies was influenced by a person's gender. A random sample was taken. The results are shown in the following table.

	Country Chosen		
	USA	**Australia**	**UK**
Male	55	26	40
Female	25	31	41

A χ^2 test was performed at the 1% significance level.
The critical value for this test is 9.210.

a. State the null hypothesis. [1]

b. Write down the number of degrees of freedom. [1]

c. Write down [2]

(i) the χ^2 statistic;

(ii) the associated *p*-value.

d. State, giving a reason, whether the null hypothesis should be accepted. [2]

Markscheme

a. Country chosen and gender are independent. ***(A1)*** ***(C1)***

Notes: Accept there is no association between country chosen and gender.

Do not accept "not related" or "not correlated" or "influenced".

[1 mark]

b. 2 ***(A1)*** ***(C1)***

[1 mark]

c. (i) 9.17 (9.16988…) ***(A1)***

Notes: Accept 9.169.

(ii) 0.0102 (0.0102043…) ***(A1)*** ***(C2)***

Notes: Award ***(A1)*** for 0.010, but ***(A0)*** for 0.01.

[2 marks]

d. Since $0.0102 > 0.01$, we accept the null hypothesis. ***(R1)(A1)*****(ft)**

OR

Since $9.17 < 9.210$, we accept the null hypothesis. ***(R1)(A1)*****(ft)** ***(C2)***

Notes: To award ***(R1)*** there should be value(s) given in part (c). If a value is given in (c), we do not need it explicitly stated again in (d).

It is sufficient to state a correct comparison.

e.g. p-value $>$ significance level **OR** $\chi^2_{\text{calc}} <$ critical value

Do not award ***(R0)(A1)***. Follow through from part (c).

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

The number of calories a person burns during a walk depends on the time they spend walking. The table below shows the number of calories, y, burned by a person in relation to the time they spend walking, x, in minutes.

Time spent walking (x) (minutes)	10	15	20	25	30
Calories (y)	90	125	200	300	375

a. Use your graphic display calculator to write down the equation of the regression line for y on x in the form $y = ax + b$. [2]

b. Use your equation to estimate the number of calories that a person will burn during a 17 minute walk. [2]

c. State whether your answer to part (b) is reliable. Give a reason for your answer. [2]

Markscheme

a. $y = 14.9x - 80$ ***(A1)(A1)*** ***(C2)***

Notes: Award ***(A1)*** for $14.9x$ and ***(A1)*** for -80. Award at most ***(A0)(A1)*** if not given in the form of an equation.

[2 marks]

b. $14.9 \times 17 - 80$ ***(M1)***

Note: Award ***(M1)*** for substitution in their equation from part (a).

173.3 calories ***(A1)*****(ft)** ***(C2)***

Note: Accept 173 and 170 even if no working is seen.

[2 marks]

c. Reliable. 17 min is in the range of given values for x **or** correlation coefficient (r) is 0.989… ***(A1)(R1)*** ***(C2)***

Notes: Do not award ***(A1)(R0)***. Alternative acceptable reasons using correlation:

Correlation coefficient close to 1

Strong positive correlation

Strong linear correlation

Strong positive association between the variables

Strong relationship between the variables.

Examiners report

a. This question was an opportunity for candidates to show effective use of the GDC and many correct answers were seen in parts (a) and (b).

b. This question was an opportunity for candidates to show effective use of the GDC and many correct answers were seen in parts (a) and (b).

c. Part (c) was unusual in that questions of this nature, in the past, have focused on values outside of the range of given values of x. For correct reasoning, candidates were required to identify that 17 was in the range of values given for x in the table. Identifying that 17 was between 15 and 20 minutes was sufficient whereas identifying that 173 was between 125 and 200 was clearly not sufficient. Alternative reasons which focused on the correlation coefficient being either strong positive or close to one were seen and were accepted.

The mass of a certain type of Chilean corncob follows a normal distribution with a mean of 400 grams and a standard deviation of 50 grams.

A farmer labels one of these corncobs as premium if its mass is greater than a grams. 25% of these corncobs are labelled as premium.

a. Write down the probability that the mass of one of these corncobs is greater than 400 grams. [1]

b. Find the value of a. [2]

c. Estimate the interquartile range of the distribution. [3]

Markscheme

a. $0.5 \left(50\%, \frac{1}{2}\right)$ ***(A1)*** ***(C1)***

[1 mark]

b. $\mathrm{P}(X > a) = 0.25$ **OR** $\mathrm{P}(X < a) = 0.75$ ***(M1)***

Note: Award ***(M1)*** for a sketch of approximate normal curve with a vertical line drawn to the right of the mean with the area to the right of this line shaded.

$a = 434\ (\mathrm{g})\ (433.724\ldots\ (\mathrm{g}))$ ***(A1)*** ***(C2)***

[2 marks]

c. $33.7244\ldots \times 2$ ***(A1)*(ft)*(M1)***

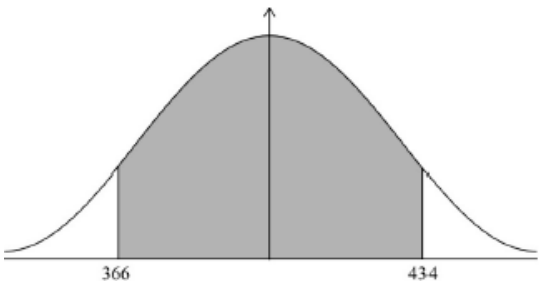
Note: Award ***(A1)*(ft)*** for $33.7244\ldots$ $(\text{or } 433.7244\ldots - 400)$ seen, award ***(M1)*** for multiplying their 33.7244… by 2. Follow through from their answer to part (b).

OR

$434 - 366.275\ldots$ ***(A1)*(ft)*(M1)***

Note: Award ***(A1)*(ft)*** for their $366.275\ldots$ (366) seen, ***(M1)*** for difference between their answer to (b) and their 366.

OR



(A1)*(ft)*(M1)

Note: Award ***(A1)*(ft)*** for their $366.275\ldots$ (366) seen. Award ***(M1)*** for correct symmetrical region indicated on labelled normal curve.

67.4 (g) ***(A1)*(ft)*** ***(C3)***

Note: Accept an answer of 68 from use of rounded values. Follow through from part (b).

[3 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

A group of 100 students gave the following responses to the question of how they get to school.

	Walk	Public transport	Car	Bicycle	Total
Female	18	13	14	3	48
Male	9	17	10	16	52
Total	27	30	24	19	100

A χ^2 test for independence was conducted at the 5% significance level. The null hypothesis was defined as

H_0: Method of getting to school is independent of gender.

a. Find the expected frequency for the females who use public transport to get to school. [2]

b. Find the χ^2 statistic. [2]

c. The χ^2 critical value is 7.815 at the 5% significance level. [2]

State whether or not the null hypothesis is accepted. Give a reason for your answer.

Markscheme

a. $\frac{30}{100} \times \frac{48}{100} \times 100$ **OR** $\frac{30 \times 48}{100}$ ***(M1)***

Note: Award ***(M1)*** for correct substitution into correct formula.

$= 14.4 \left(\frac{72}{5}\right)$ ***(A1)*** ***(C2)***

[2 marks]

b. $13.0\ (12.9554\ldots)$ ***(A2)*** ***(C2)***

Note: Award ***(A1)(A0)*** for 12.9.

[2 marks]

c. the null hypothesis is not accepted ***(A1)*****(ft)**

$\chi^2_{calc} > \chi^2_{crit}$ **OR** $13.0 > 7.82$ ***(R1)***

OR

the null hypothesis is not accepted ***(A1)*****(ft)**

p-value (0.0047) $(0.00473391\ldots) < 0.05$ ***(R1)*** ***(C2)***

Notes: Follow through from their answer to part (b).

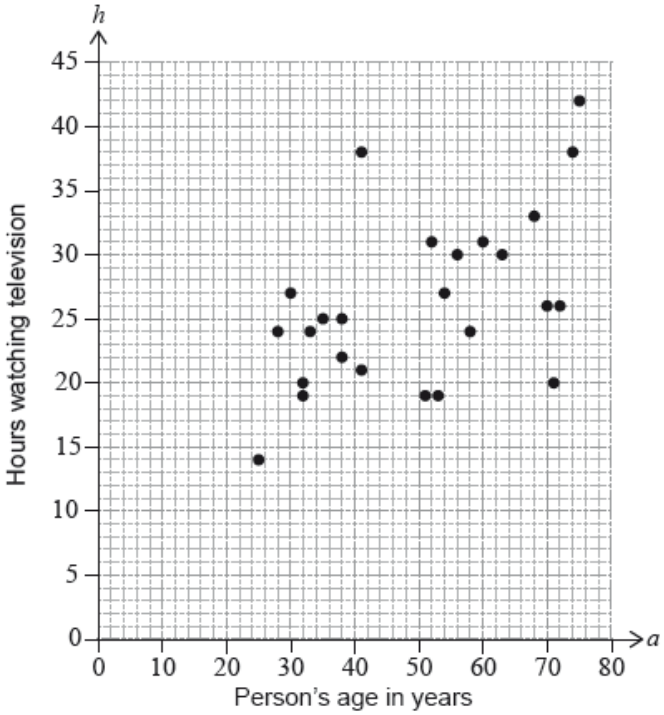
Do not award ***(A1)*****(ft)*****(R0)***.

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

A survey was carried out to investigate the relationship between a person's age in years (a) and the number of hours they watch television per week (h). The scatter diagram represents the results of the survey.



The mean age of the people surveyed was 50.

For these results, the equation of the regression line h on a is $h = 0.22a + 15$.

a. Find the mean number of hours that the people surveyed watch television per week. [2]

b. Draw the regression line on the scatter diagram. [2]

c. By placing a tick (✔) in the correct box, determine which of the following statements is true: [1]

The correlation between h and a is positive.	
The correlation between h and a is negative.	
There is no correlation between h and a.	

d. Diogo is 18 years old. Give a reason why the regression line should not be used to estimate the number of hours Diogo watches television per week. [1]

Markscheme

a. $0.22(50) + 15$ ***(M1)***

Note: Award ***(M1)*** for correct substitution of 50 into equation of the regression line.

$(=) 26$ ***(A1)*** ***(C2)***

OR

$\frac{655}{25}$ ***(M1)***

Note: Award ***(M1)*** for correctly summing the h values of the points, and dividing by 25.

$(=) 26.2$ ***(A1)*** ***(C2)***

[2 marks]

b. line through $(50,\ 26 \pm 1)$ and $(0,\ 15)$ ***(A1)*(ft)*(A1)*** ***(C2)***

Note: Award ***(A1)*(ft)** for a straight line through (50, their $\bar{h}$), and ***(A1)*** for the line intercepting the y-axis at $(0,\ 15)$; this may need to be extrapolated. Follow through from part (a). Award at most ***(A0)(A1)*** if the line is not drawn with a ruler.

[2 marks]

c.

The correlation between h and a is positive.	✓
The correlation between h and a is negative.	
There is no correlation between h and a.	

(A1) (C1)

Note: Award ***(A0)*** if more than one tick (✔) is seen.

[1 mark]

d. 18 is less than the lowest age in the survey **OR** extrapolation. ***(A1)*** ***(C1)***

Note: Accept equivalent statements.

[1 mark]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

A market researcher surveyed men and women about their preferred holiday destination. The holiday destinations were Antigua, Barbados, Cuba, Guadeloupe and Jamaica. A χ^2 test for independence was conducted at the 5 % significance level.

The χ^2 calculated value was found to be 8.73.

a. Write down the null hypothesis. [1]

b. Find the number of degrees of freedom for this test. [2]

c. Write down the critical value for this test. [1]

d. State the conclusion of this test. Give a reason for your decision. [2]

Markscheme

a. Holiday destination is independent of gender. ***(A1)*** ***(C1)***

Note: Accept gender is independent of holiday destination, accept *"not associated"*, do not accept *"not correlated"*.

b. $(2-1)(5-1)$ ***(M1)***

Note: Award ***(M1)*** for correct substitution in the correct formula.

$= 4$ ***(A1)*** ***(C2)***

c. 9.488 ***(A1)*****(ft)** ***(C1)***

Notes: Follow through from their answer to part (b). Accept 9.49.

d. Accept the null hypothesis **or** Accept H_0. ***(A1)*****(ft)**

Note: Accept gender is independent of holiday destination.

$\chi^2_{(\text{calc})} < \chi^2_{(\text{crit})}$ **or** 8.73 < 9.488 ***(R1)*** ***(C2)***

Notes: Do not award ***(A1)(R0)***. Follow through from their answer in part (c).

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

A survey investigated the relationship between the number of cleaners, n, and the amount of time, t, it takes them to clean a school.

Number of cleaners, n	Time, t (minutes)
1	193
2	172
3	118
5	112
6	87

a. Use your graphic display calculator to write down the equation of the regression line t on n. [2]

b. Write down the value of the Pearson's product–moment correlation coefficient, r. [2]

c. Use your regression equation to find the amount of time 4 cleaners take to clean the school. [2]

Markscheme

a. $t = -20.1n + 205$

$t = (-20.1046\ldots)n + (204.755\ldots)$ ***(A1)(A1)*** ***(C2)***

Notes: Award ***(A1)*** for -20.1 and 205 seen,

(A1) for an equation involving t and n.

[2 marks]

b. $-0.941(-0.941366\ldots)$ ***(A2)*** ***(C2)***

Notes: Award ***(A0)(A1)*** for $+0.941$.

[2 marks]

c. $-20.1046\ldots \times 4 + 204.755\ldots$ ***(M1)***

Note: Award ***(M1)*** for substitution into their regression equation.

124 (minutes) $(124.337\ldots)$ ***(A1)*****(ft)** ***(C2)***

Notes: Follow through from their regression equation found in part (a). Accept 125 (minutes) (124.6).

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

The marks obtained by 8 candidates in Physics and Chemistry tests are given below.

Physics (x)	6	8	10	11	10	5	4	12
Chemistry (y)	8	11	14	13	11	7	5	15

a. Write down the product moment correlation coefficient, r. [1]

b. Write down, in the form $y = mx + c$, the equation of the regression line y on x for the 8 candidates. [2]

c. A ninth candidate obtained a score of 7 in the Physics test but was absent for the Chemistry test. [2]

Use your answer to (b) to estimate the score he would have obtained on the Chemistry test.

d. Give a reason why it is valid to use this regression line to estimate the score on the Chemistry test in part (c). [1]

Markscheme

a. 0.965 ***(A1)*** ***(C1)***

[1 mark]

b. $y = 1.15x + 0.976$ ***(A1)(A1)*** ***(C2)***

Note: ***(A1)*** for $1.15x$. ***(A1)*** for $+0.976$.

[2 marks]

c. $y = 1.15(7) + 0.976$ ***(M1)***

$\text{Chemistry} = 9.03$ *(accept 9)* ***(A1)*****(ft)** ***(C2)***

Note: Follow through from candidate's answer to (b) even if no working is seen. Award ***(A2)*****(ft)**.

[2 marks]

d. the correlation coefficient is close to 1

OR strongly correlated variables

OR 7 lies within the range of physics marks. ***(R1)*** ***(C1)***

[1 mark]

Examiners report

a. The level of accuracy required by the paper was often ignored in this question.

(a) Some candidates are unable to recover r from a reset calculator.

b. The level of accuracy required by the paper was often ignored in this question.

c. The level of accuracy required by the paper was often ignored in this question.

(c) Many candidates seem to be unaware when it is appropriate to use a regression line.

d. The level of accuracy required by the paper was often ignored in this question.

Applicants for a job had to complete a mathematics test. The time they took to complete the test is normally distributed with a mean of 53 minutes and a standard deviation of 16.3. One of the applicants is chosen at random.

For 11% of the applicants it took longer than k minutes to complete the test.

There were 400 applicants for the job.

a. Find the probability that this applicant took at least 40 minutes to complete the test. [2]

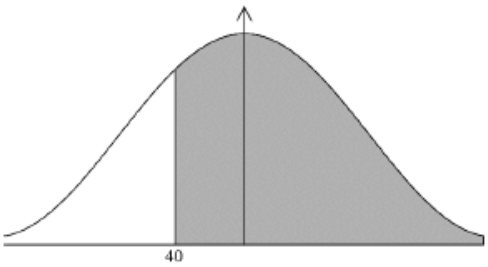
b. Find the value of k. [2]

c. Estimate the number of applicants who completed the test in less than 25 minutes. [2]

Markscheme

a. 0.787 (0.787433…, 78.7%) ***(M1)(A1)*** ***(C2)***

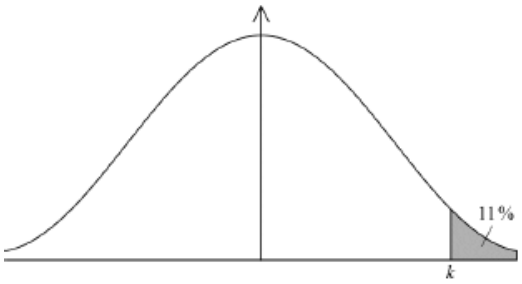
Note: Award ***(M1)*** for a correct probability statement, $\mathrm{P}(X > 40)$, or a correctly shaded normal distribution graph.



[2 marks]

b. 73.0 (minutes) (72.9924…) ***(M1)(A1)*** ***(C2)***

Note: Award ***(M1)*** for a correct probability statement, $\mathrm{P}(X > k) = 0.11$, or a correctly shaded normal distribution graph.



[2 marks]

c. $0.0423433\ldots \times 400$ ***(M1)***

Note: Award ***(M1)*** for multiplying a probability by 400. Do not award ***(M1)*** for 0.11×400.

Use of a lower bound less than zero gives a probability of 0.0429172….

$= 16$ ***(A1)*** ***(C2)***

Notes: Accept a final answer of 17. Do not accept a final answer of 18. Accept a non-integer final answer either 16.9 (16.9373…) from use of lower bound zero or 17.2 (17.1669…) from use of the default lower bound of -10^{99}.

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

Tania wishes to see whether there is any correlation between a person's age and the number of objects on a tray which could be remembered after looking at them for a certain time.

She obtains the following table of results.

Age (x years)	15	21	36	40	44	55
Number of objects remembered (y)	17	20	15	16	17	12

a. Use your graphic display calculator to find the equation of the regression line of y on x. [2]

b. Use your equation to estimate the number of objects remembered by a person aged 28 years. [1]

c. Use your graphic display calculator to find the correlation coefficient r. [1]

d. Comment on your value for r. [2]

Markscheme

a. $a = -0.134, b = 20.9$ ***(A1)***

$y = 20.9 - 0.134x$ ***(A1)*** ***(C2)***

[2 marks]

b. 17 objects ***(A1)*(ft)**

accept only 17 ***(C1)***

[1 mark]

c. $r = -0.756$ ***(A1)*** ***(C1)***

[1 mark]

d. negative and moderately strong ***(A1)*(ft)*(A1)*(ft)** ***(C2)***

[2 marks]

Examiners report

a. This question expected the candidates to use their GDC to find the equation of the regression line and to find the correlation coefficient and this was stated in the question parts. However, there were a number of candidates from specific schools who had not been taught to do this and they tried to find the equation from the formula in the information booklet. This would have wasted time and they did not manage to find the correct answer. This was also the case for the correlation coefficient. However, most candidates who found a correlation coefficient managed to comment on it correctly.

b. This question expected the candidates to use their GDC to find the equation of the regression line and to find the correlation coefficient and this was stated in the question parts. However, there were a number of candidates from specific schools who had not been taught to do this and they tried to find the equation from the formula in the information booklet. This would have wasted time and they did not manage to find the correct answer. This was also the case for the correlation coefficient. However, most candidates who found a correlation coefficient managed to comment on it correctly.

Many candidates did not give a whole number answer for the number of objects in part (b).

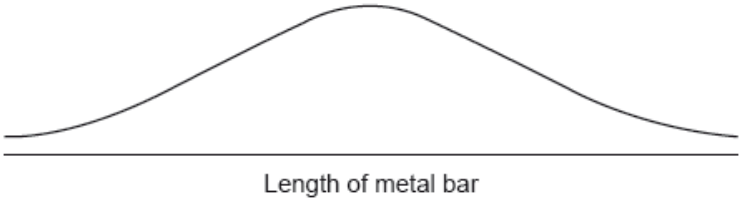
c. This question expected the candidates to use their GDC to find the equation of the regression line and to find the correlation coefficient and this was stated in the question parts. However, there were a number of candidates from specific schools who had not been taught to do this and they tried to find the equation from the formula in the information booklet. This would have wasted time and they did not manage to find the correct answer. This was also the case for the correlation coefficient. However, most candidates who found a correlation coefficient managed to comment on it correctly.

d. This question expected the candidates to use their GDC to find the equation of the regression line and to find the correlation coefficient and this was stated in the question parts. However, there were a number of candidates from specific schools who had not been taught to do this and they tried to find the equation from the formula in the information booklet. This would have wasted time and they did not manage to find the correct

answer. This was also the case for the correlation coefficient. However, most candidates who found a correlation coefficient managed to comment on it correctly.

A factory makes metal bars. Their lengths are assumed to be normally distributed with a mean of 180 cm and a standard deviation of 5 cm.

a. On the following diagram, shade the region representing the probability that a metal bar, chosen at random, will have a length less than 175 cm. [2]



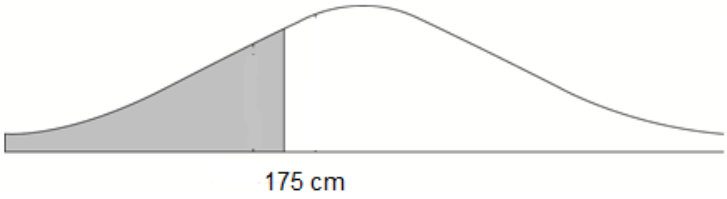
b. A metal bar is chosen at random. [4]

(i) The probability that the length of the metal bar is less than 175 cm is equal to the probability that the length is greater than h cm. Write down the value of h.

(ii) Find the probability that the length of the metal bar is greater than one standard deviation above the mean.

Markscheme

a.



(A1)(M1) ***(C2)***

Notes: Award ***(A1)*** for the vertical line labelled as $175\ (\text{cm})$.

Award ***(M1)*** for a vertical line drawn to the left of the mean with the area to the left of this line shaded.

Accept $(-)1$ sd marked on the diagram for 175 (provided line is to the left of the mean).

b. (i) $185\ (\text{cm})$ ***(A1)(C1)***

(ii) $\text{P}(\text{length} > 185)$ ***(A1)(M1)***

Note: Award ***(A1)*** for the vertical line labelled as $185\ (\text{cm})$.

Award ***(M1)*** for a vertical line drawn to the right of the mean with the area to the right of this line shaded.

Accept 1 sd marked on the diagram for 185 (provided line is to the right of the mean).

$= 0.159 \quad (0.158655\ldots)$ ***(A1)*** ***(C3)***

Examiners report

[N/A]

b. [N/A]

Minta surveyed students from her school about their preferred morning snack from a choice of an apple, a fruit salad or a smoothie.

She surveyed 350 students, of whom 210 are female.

She performed a χ^2 test at the 5% significance level to determine whether there is a relationship between the choice of morning snack and gender.

a. State Minta's null hypothesis. [1]

b. State the number of degrees of freedom. [1]

c. 150 students showed a preference for a smoothie. [2]

Calculate the expected number of female students who chose a smoothie.

d. Minta found that the calculated value of the χ^2 test was 3.576. The critical value at the 5% significance level is 5.99. [2]

State Minta's conclusion. Give a reason for your answer.

Markscheme

a. H_0: Choice of morning snack is independent of (not dependent on) gender. ***(A1)*** ***(C1)***

Note: Accept there is "no association" between snack chosen and gender.

Do not accept "not related" or "not correlated" or "influenced".

b. 2 ***(A1)*** ***(C1)***

c. $\frac{210\times150}{350}$ ***(M1)***

Note: Award ***(M1)*** for correct substitution in the correct formula.

$= 90$ ***(A1)*** ***(C2)***

d. Null hypothesis is accepted (not rejected). ***(A1)***

OR

Choice of morning snack is independent of gender ***(A1)***

$3.576 < 5.99 \quad \mathbf{OR} \quad \chi^2_{\text{calc}} < \chi^2_{\text{crit}}$ ***(R1)*** ***(C2)***

Note: Do not award ***(A1)(R0)***.

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

In a school, students in grades 9 to 12 were asked to select their preferred drink. The choices were milk, juice and water. The data obtained are organized in the following table.

	Milk	Juice	Water	**Total**
Grade 9	25	34	15	**74**
Grade 10	31	x	13	**74**
Grade 11	18	35	17	**70**
Grade 12	9	36	26	**71**
Total	**83**	**135**	**71**	**289**

A χ^2 test is carried out at the 5% significance level with hypotheses:

H_0: the preferred drink is independent of the grade
H_1: the preferred drink is not independent of the grade

The χ^2 critical value for this test is 12.6.

a. Write down the value of x. [1]

b. Write down the number of degrees of freedom for this test. [1]

c. Use your graphic display calculator to find the χ^2 statistic for this test. [2]

d. State the conclusion for this test. Give a reason for your answer. [2]

Markscheme

a. 30 ***(A1)*** ***(C1)***

[1 mark]

b. 6 ***(A1)*** ***(C1)***

[1 mark]

c. $19.0\ (18.9640)\ldots$ ***(A2)*****(ft)** ***(C2)***

Note: Follow through from part (a).

Award ***(A1)*** for truncation to 18.9.

[2 marks]

d. reject (do not accept) H_0 **OR** accept H_1 ***(A1)*****(ft)**

Note: Can be written in words.

$19.0\ (18.9640\ldots) > 12.6$ ***(R1)*** ***(C2)***

Note: Accept "$\chi^2_{calc} > \chi^2_{crit}$" for the ***(R1)*** provided their χ^2_{calc} value is explicitly seen in their part (c).

OR

$(p =)\ 0.00422\ <(\text{significance level}\ =)\ 0.05$ ***(R1)*** ***(C2)***

Note: Do not award ***(R0)(A1)*****(ft)**. Follow through from part (c). Numerical comparison must be seen to award the ***(R1)***.

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

A shop keeper recorded daily sales s of ice cream along with the daily maximum temperature t °C. The results for one week are shown below.

t	29	31	34	23	19	20	27
s	104	92	112	48	56	72	66

a. Write down the equation of the regression line for s on t. [3]

b. Use your equation to predict the ice cream sales on a day when the maximum temperature is 24 °C. Give your answer correct to the nearest whole number. [3]

Markscheme

a. $s = 3.56\,t - 14.6$ ***(A1)(A1)(A1)*** ***(C3)***

Notes: Award ***(A1)*** for 3.56.

(A1) for –14.6.

(A1) for *s* and *t*.

[3 marks]

b. $s = 3.56 \times 24 - 14.6$ ***(M1)***

$= 70.84$ (70.9) ***(A1)*****(ft)**

= 71 ice creams ***(A1)*****(ft)** ***(C3)***

Note: (ft) from candidates answer to (a).

Note: The last ***(A1)*** is for specified accuracy, **(ft)** from their answer.

The ***(AP)*** for the paper is not applied here.

[3 marks]

Examiners report

a. Some candidates attempted to find the equation by hand, generally without success. Those who used their calculator could quickly find the equation and use it to find the number of ice cream sales. A significant number of candidates lost one mark for writing the equation with y and x rather than s and t. A lesser number lost the accuracy mark for an integral number of ice-creams.

b. Some candidates attempted to find the equation by hand, generally without success. Those who used their calculator could quickly find the equation and use it to find the number of ice cream sales. A significant number of candidates lost one mark for writing the equation with y and x rather than s and t. A lesser number lost the accuracy mark for an integral number of ice-creams.

The scores obtained by five candidates in Mathematics and Physics examinations are given below.

Mathematics (x)	62	84	47	55	32
Physics (y)	80	91	44	48	53

a. Write down the correlation coefficient, r , for the examination scores. [2]

b. Write down the equation of the regression line, y on x , for the examination scores of the five candidates. [2]

c. A sixth candidate scored 72 in the Mathematics examination. Use the regression line, y on x, to estimate his score on the Physics examination. [2]

Markscheme

a. $r = 0.814\ (0.813745...)$ ***(A2)*** ***(C2)***

[2 marks]

b. $y = 0.888x + 13.5 \quad (y = 0.887686\ldots x + 13.4895\ldots)$ ***(A1)(A1)***

Note: Award ***(A1)*** for $0.888x$, ***(A1)*** for 13.5. If the answer is not in the form of an equation award ***(A1)(A0)***.

OR

$y - 63.2 = 0.888(x - 56)$ ***(A1)(A1)*** ***(C2)***

Note: Award ***(A1)*** for 0.888, ***(A1)*** for the correct means, $\bar{x}$ and $\bar{y}$ used.

[2 marks]

c. $y = 0.887686\ldots(72) + 13.4895\ldots$ ***(M1)***

Note: Award ***(M1)*** for 72 substituted into their equation of the regression line.

$= 77 \quad (77.4028\ldots)$ ***(A1)*****(ft)** ***(C2)***

Note: Accept a correct **(ft)** integer value or a decimal value which would round to the required 3 sf answer **(ft)**. Follow through from their equation in part (b).

[2 marks]

Examiners report

a. Able candidates scored very well on this question showing good use of their GDC and marks in excess of 4 were achieved by a large majority of these candidates. It was clear however, from some responses, that either the topic had not been taught or had been poorly understood by candidates and few, if any, marks were achieved by such candidates.

b. Able candidates scored very well on this question showing good use of their GDC and marks in excess of 4 were achieved by a large majority of these candidates. It was clear however, from some responses, that either the topic had not been taught or had been poorly understood by candidates and few, if any, marks were achieved by such candidates. In part (b), many candidates quoted a correct regression line equation using $\bar{x}, \bar{y}, s_{xy}$ and s_{x^2} but then seemed to be at a loss as to what to do with it.

c. Able candidates scored very well on this question showing good use of their GDC and marks in excess of 4 were achieved by a large majority of these candidates. It was clear however, from some responses, that either the topic had not been taught or had been poorly understood by candidates and few, if any, marks were achieved by such candidates. In part (b), many candidates quoted a correct regression line equation using $\bar{x}, \bar{y}, s_{xy}$ and s_{x^2} but then seemed to be at a loss as to what to do with it.

Identical mosquito traps are placed at different distances from a lake. On one day the number of mosquitoes caught in 10 of the traps is recorded.

Distance, $\mathbf{m}(x)$	8	15	22	30	34	45	50	60	74	82
Number of mosquitoes (y)	78	75	72	67	66	59	59	53	48	43

It is believed the number of mosquitoes caught varies linearly with the distance, in metres, of the trap from the lake.

a. Find [4]

(i) Pearson's product–moment correlation coefficient, r;

(ii) the equation of the regression line y on x.

b. Use the equation of the regression line y on x to estimate the number of mosquitoes caught in a trap that is 28 m from the lake. [2]

Markscheme

a. (i) $-0.998 \quad (-0.997770\ldots)$ ***(A2)***

Note: Award ***(A0)(A1)*** for $0.998(0.997770\ldots)$.

Award ***(A1)(A0)*** for -0.997.

(ii) $y = -0.470x + 81.7 \quad (y = -0.469713\ldots x + 81.7279\ldots)$ ***(A1)(A1)*** ***(C4)***

Note: Award a maximum of ***(A0)(A1)*** if the answer is not an equation.

b. $-0.469713\ldots(28) + 81.7279$ ***(M1)***

Note: Award ***(M1)*** for correct substitution of 28 into their equation of regression line.

$= 68.6\ (\text{mosquitoes}) \quad (68.5759\ldots)$ ***(A1)*****(ft)** ***(C2)***

Note: Accept 68 or 69 or $68.5(4)$ from use of 3 sf values.

Follow through from part (a)(ii).

Examiners report

a. In part (a)(i), the majority of candidates knew how to calculate Pearson's correlation coefficient using their GDC. The most common errors were incorrect rounding and omitting the – sign. In part (a)(ii) many candidates correctly found the equation of the regression line, again with rounding errors being the most common. A very common error was to use the second list as the frequency for the statistics.

b. In part (b) substitution of 28 in the regression line was done correctly by many candidates. Candidates seemed to be well prepared for this type of question.

---

200 people of different ages were asked to choose their favourite type of music from the choices Popular, Country and Western and Heavy Metal. The results are shown in the table below.

Age/Music choice	Popular	Country and Western	Heavy Metal	Totals
11 – 25	35	5	50	90
26 – 40	30	10	20	60
41 – 60	20	25	5	50
Totals	85	40	75	200

It was decided to perform a chi-squared test for independence at the 5% level on the data.

a. Write down the null hypothesis. [1]

b. Write down the number of degrees of freedom. [1]

c. Write down the chi-squared value. [2]

d. State whether or not you will reject the null hypothesis, giving a clear reason for your answer. [2]

Markscheme

a. Choice of music is independent of age. ***(A1)*** ***(C1)***

[1 mark]

b. $(3-1)(3-1)$

$=4$ ***(A1)*** ***(C1)***

[1 mark]

c. $\chi^2 = 51.6$ ***(A2)*** ***(C2)***

Note: 52 is an accuracy penalty ***(A1)(A0)(AP)***.

[2 marks]

d. p-value < 0.05 for 5% level of significance ***(R1)*****(ft)**

or $51.6 > \chi^2_{crit}$

Reject the null hypothesis *(do not accept the null hypothesis)*. ***(A1)*****(ft)** ***(C2)***

Note: Do not award ***(R0)(A1)***.

[2 marks]

Examiners report

a. Candidates either gained good marks for this question or almost no marks depending on their preparation. It was obvious that some schools had omitted this from their programme. Candidates generally gave a reason for their conclusion in part (d) though some compared the chi-squared value with the p-value, resulting in the loss of both marks.

b. Candidates either gained good marks for this question or almost no marks depending on their preparation. It was obvious that some schools had omitted this from their programme. Candidates generally gave a reason for their conclusion in part (d) though some compared the chi-squared value with the p-value, resulting in the loss of both marks.

c. Candidates either gained good marks for this question or almost no marks depending on their preparation. It was obvious that some schools had omitted this from their programme. Candidates generally gave a reason for their conclusion in part (d) though some compared the chi-squared value with the p-value, resulting in the loss of both marks.

d. Candidates either gained good marks for this question or almost no marks depending on their preparation. It was obvious that some schools had omitted this from their programme. Candidates generally gave a reason for their conclusion in part (d) though some compared the chi-squared value with the p-value, resulting in the loss of both marks.

A questionnaire was given to all members of a school community to find out which drink was the most popular to have with breakfast. The results are given in the table below, classified by age.

	Hot Chocolate	Tea	Coffee	Milk
Children aged 12 years and less	55	10	1	34
Teenagers aged from 13 to 19 years	25	35	20	10
Adults aged 20 years and over	20	40	79	6

A χ^2 test was conducted to decide whether the type of drink was independent of age.

a. Find the number of degrees of freedom for the χ^2 test. [2]

b. Write down the null hypothesis for the χ^2 test. [1]

d. The critical value for the χ^2 test at the 5% significance level is 12.59. The χ^2 test statistic is calculated to be 146 with a *p*-value of 6.62×10^{-29} [2]

(both numbers given correct to 3 significant figures).

Write down the conclusion reached at the 5 % significance level. Give a clear reason for your answer.

Markscheme

a. $(3-1)(4-1)$ ***(M1)***

= 6 ***(A1)*** ***(C2)***

[2 marks]

b. The preferred type of drink is **independent** of age. ***(A1)*** ***(C1)***

Note: For independent accept “not associated” but do not accept “not related” or “not correlated”

[1 mark]

d. Reject null hypothesis as critical value $< \chi^2_{calc}$ ***(A1)(R1)*****(ft)**

Note: (ft) from their value in (c).

OR

Reject null hypothesis as *p*-value < 0.05 ***(A1)(R1)*** ***(C2)***

Notes: Do not award ***(A1)(R0)***.

Award the ***(R1)*** for comparison of correct values.

[2 marks]

Examiners report

a. This question caused significant difficulties for many candidates. It seemed, from the responses, that the purpose of the test is not well understood, even if its procedure on the GDC can be performed.

The test is one of "independence" and it should be stressed to candidates that it is this which is key in stating the hypotheses. Improper terminology, most notably, "not correlated" is not acceptable.

b. This question caused significant difficulties for many candidates. It seemed, from the responses, that the purpose of the test is not well understood, even if its procedure on the GDC can be performed.

d. This question caused significant difficulties for many candidates. It seemed, from the responses, that the purpose of the test is not well understood, even if its procedure on the GDC can be performed.

Many candidates did not know the correct figures to compare in order to arrive at the decision. Others gave no reason at all.

A hospital collected data from 1000 patients in four hospital wards to review the quality of its healthcare. The data, showing the number of patients who became infected during their stay in hospital, was recorded in the following table.

	Ward				
	Pasteur	**Nightingale**	**Jenner**	**Fleming**	**Total**
Patients infected	44	27	13	16	100
Patients not infected	281	303	182	134	900
Total	325	330	195	150	1000

A χ^2-test was performed at the 5% significance level.

The critical value for this test is 7.815.

The null hypothesis for the test is

H_0: Becoming infected during a stay in the hospital is independent of the ward.

a. Find the expected frequency of the patients who became infected whilst in Nightingale ward. [2]

b. For this test, write down the χ^2 statistic. [2]

c. State, giving a reason, whether the null hypothesis should be rejected. [2]

Markscheme

a. $\frac{100}{1000} \times \frac{330}{1000} \times 1000$ **OR** $\frac{100 \times 330}{1000}$ ***(M1)***

$= 33$ ***(A1)*** ***(C2)***

[2 marks]

b. 8.21 (8.21497…) ***(A2)*** ***(C2)***

[2 marks]

c. H_0 should be rejected ***(A1)*****(ft)**

$7.815 < 8.21$ **OR** (the p-value) $0.041771 < 0.05$ ***(R1)*** ***(C2)***

Note: Follow through from part (b). Do not award ***(A1)(R0)***.

Award ***(A1)*****(ft)** for "H_0 should be rejected" **OR** "Becoming infected during a stay in hospital is **not** independent of (is dependent on **OR** associated with) the ward". Accept "Do not accept H_0" **OR** "YES". Do not accept "Becoming infected during a stay in hospital is correlated (related **OR** linked) with the ward."

Award ***(R1)*** for comparison of their χ^2 statistic value from part (b) with the critical value **OR** a comparison of p-value with 0.05.

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

The daily January temperature of Cairns is normally distributed with a mean of 34°C and a standard deviation of 3.

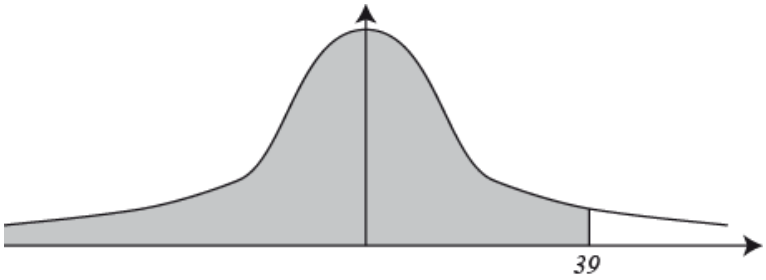
a. Calculate the probability that the temperature on a randomly chosen day in January is less than 39°C. [2]

b. Calculate the expected number of days in January that the temperature will be more than 39°C. [2]

c. On a randomly chosen day in January, the probability that the temperature is above T °C is 0.7. [2]

Find the value of T.

Markscheme

a. 

(M1)

Note: Award ***(M1)*** for correctly shaded area.

$0.952\ (95.2\%,\ 0.952209\ldots)$ ***(A1)*** ***(C2)***

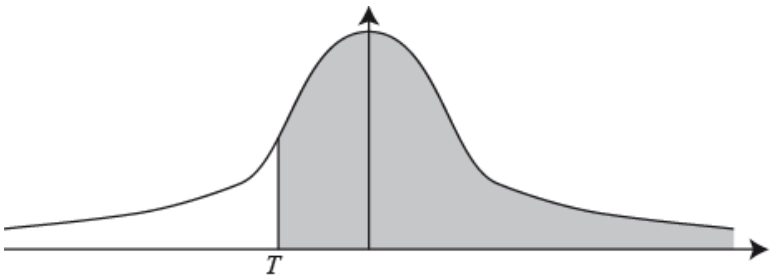
b. $31 \times (1 - 0.952209)$ ***(M1)***

Note: Award ***(M1)*** for multiplying 31 by $(1 -$ their answer to part (a)$)$.

$= 1.48\ (1.48150\ldots)$ ***(A1)*****(ft)** ***(C2)***

Note: Follow through from part (a).

c.



(M1)

Note: Award ***(M1)*** for correctly shaded area.

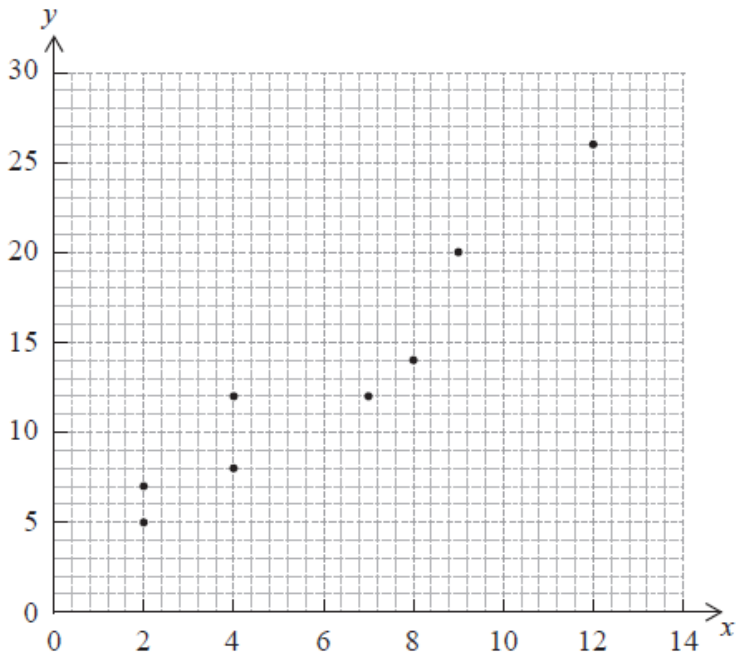
$32.4\ (32.4267\ldots)$ ***(A1)*** ***(C2)***

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

Consider the following set of data which is plotted on the scatter diagram below.

x	2	4	7	12	4	8	9	2
y	5	8	12	26	12	14	20	7



a. Write down the coordinates of the mean point $(\bar{x}, \bar{y})$. [2]

b. Write down the value of r, the Pearson's product-moment correlation coefficient for this set of data. [2]

c. Draw the regression line for y on x on the set of axes above. [2]

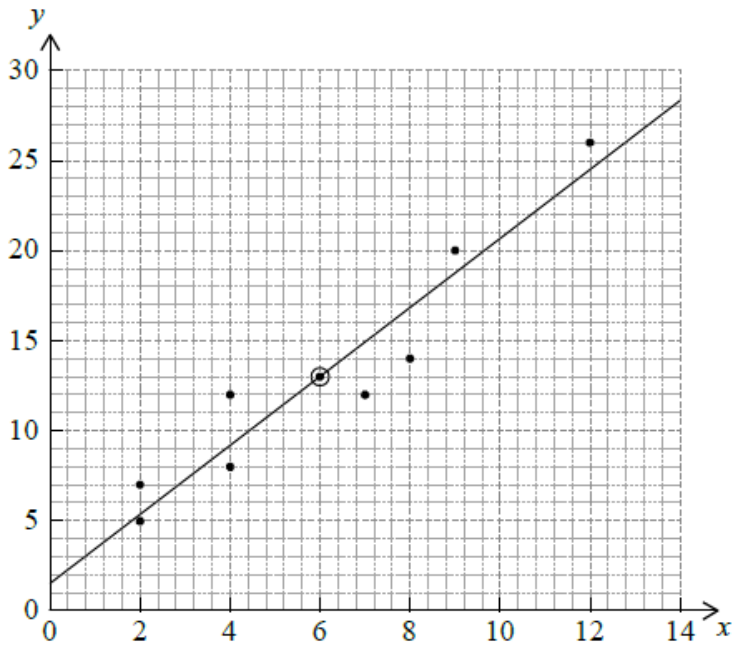
Markscheme

a. $(6, 13)$ ***(A1)(A1)*** ***(C2)***

b. $0.952\ (0.95202\ldots)$ ***(A2)*** ***(C2)***

Note: Award ***(A0)*** for 0.9.

c.



(A1) y intercept at $y = 1.8$ (*accept between 1 and 2*)
(A1)**(ft)** line passes through their mean point ***(A1)(A1)*****(ft)** ***(C2)***

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

---

a. The manager of a travel agency surveyed 1200 travellers. She wanted to find out whether there was a relationship between a traveller's age and their preferred destination. The travellers were asked to complete the following survey. [1]

Traveller survey

My age is:

25 or younger	26–40	41–60	61 or older

My preferred destination is:

New York	Tokyo	Melbourne	Dubai	Marrakech

A χ^2 test was carried out, at the 5% significance level, on the data collected.

Write down the null hypothesis.

b. Find the number of degrees of freedom. [2]

c. The critical value of this χ^2 test is 21.026. [1]

Use this information to write down the values of the χ^2 statistic for which the null hypothesis is rejected.

d. From the travellers taking part in the survey, 285 were 61 years or older and 420 preferred Tokyo. [2]

Calculate the expected number of travellers who preferred Tokyo and were 61 years or older.

Markscheme

a. age and preferred destination are independent ***(A1) (C1)***

Note: Accept there is no association between preferred destination and age. Accept not dependent. Do not accept "not related" or "not correlated" or "influenced".

b. $(4-1)\times(5-1)$ ***(M1)***

Note: Award ***(M1)*** for 3 and 4 ($4-1$ and $5-1$) seen.

$=12$ ***(A1) (C2)***

c. $\chi^2_{calc} > 21.026$ **OR** $(21.026,\ \infty)$ **OR** $]21.026\,,\ \infty[$ ***(A1) (C1)***

Note: Do not accept $\chi^2_{calc} > \chi^2_{crit}$ without numerical value.

d. $\frac{285}{1200}\times\frac{420}{1200}\times 1200$ $\left(\frac{285\times420}{1200}\right)$ ***(M1)***

Note: Award ***(M1)*** for correct substitution into correct formula.

$=99.8$ (99.75) ***(A1) (C2)***

Examiners report

a. Question 10: χ^2 test

Without doubt, this question was the subject of most comment in the feedback from teachers; opinion was divided between those who wanted the straightforward formulaic approach that can be taught in a recipe-book manner and those who saw the critical thinking nature of the question's

intent; being able to take the data, set up the two-way table and design the test. These are necessary skills for the IA project and they should be transferrable to an examination. That said, the unfamiliar form of the question caught a number of candidates unaware, most noticeably in the statement of the critical region; which is still felt – despite the GDC's use of the p-value – to be an important concept that should be introduced to candidates and will continue to be tested. Calculating the degrees of freedom was also subject to many errors.

b. Question 10: χ^2 test Without doubt, this question was the subject of most comment in the feedback from teachers; opinion was divided between those who wanted the straightforward formulaic approach that can be taught in a recipe-book manner and those who saw the critical thinking nature of the question's intent; being able to take the data, set up the two-way table and design the test. These are necessary skills for the IA project and they should be transferrable to an examination. That said, the unfamiliar form of the question caught a number of candidates unaware, most noticeably in the statement of the critical region; which is still felt – despite the GDC's use of the p-value – to be an important concept that should be introduced to candidates and will continue to be tested. Calculating the degrees of freedom was also subject to many errors.

c. Question 10: χ^2 test Without doubt, this question was the subject of most comment in the feedback from teachers; opinion was divided between those who wanted the straightforward formulaic approach that can be taught in a recipe-book manner and those who saw the critical thinking nature of the question's intent; being able to take the data, set up the two-way table and design the test. These are necessary skills for the IA project and they should be transferrable to an examination. That said, the unfamiliar form of the question caught a number of candidates unaware, most noticeably in the statement of the critical region; which is still felt – despite the GDC's use of the p-value – to be an important concept that should be introduced to candidates and will continue to be tested. Calculating the degrees of freedom was also subject to many errors.

d. Question 10: χ^2 test Without doubt, this question was the subject of most comment in the feedback from teachers; opinion was divided between those who wanted the straightforward formulaic approach that can be taught in a recipe-book manner and those who saw the critical thinking nature of the question's intent; being able to take the data, set up the two-way table and design the test. These are necessary skills for the IA project and they should be transferrable to an examination. That said, the unfamiliar form of the question caught a number of candidates unaware, most noticeably in the statement of the critical region; which is still felt – despite the GDC's use of the p-value – to be an important concept that should be introduced to candidates and will continue to be tested. Calculating the degrees of freedom was also subject to many errors.

A scientist measures the concentration of dissolved oxygen, in milligrams per litre (y) , in a river. She takes 10 readings at different temperatures, measured in degrees Celsius (x).

The results are shown in the table.

Temperature, °C (x)	20	24	25	28	29	32	27	25	23	21
Dissolved Oxygen, mg l^{-1} (y)	10.9	9.7	9.2	7.6	7.3	6.4	7.9	8.4	9.4	9.9

It is believed that the concentration of dissolved oxygen in the river varies linearly with the temperature.

a.i. For these data, find Pearson's product-moment correlation coefficient, *r.* [2]

a.ii. For these data, find the equation of the regression line *y* on *x*. [2]

b. Using the equation of the regression line, estimate the concentration of dissolved oxygen in the river when the temperature is 18 °C. [2]

Markscheme

a.i. −0.974 (−0.973745…) ***(A2)***

Note: Award ***(A1)*** for an answer of 0.974 (minus sign omitted). Award ***(A1)*** for an answer of −0.973 (incorrect rounding).

[2 marks]

a.ii. $y = -0.365x + 17.9$ ($y = -0.365032\ldots x + 17.9418\ldots$) ***(A1)(A1) (C4)***

Note: Award ***(A1)*** for $-0.365x$, ***(A1)*** for 17.9. Award at most ***(A1)(A0)*** if not an equation or if the values are reversed (eg $y = 17.9x - 0.365$).

[2 marks]

b. $y = -0.365032\ldots \times 18 + 17.9418\ldots$ ***(M1)***

Note: Award ***(M1)*** for correctly substituting 18 into their part (a)(ii).

= 11.4 (11.3712…) ***(A1)*(ft) *(C2)***

Note: Follow through from part (a)(ii).

[2 marks]

Examiners report

a.i. [N/A]
a.ii. [N/A]
b. [N/A]

---

a. A survey was conducted among a random sample of people about their favourite TV show. People were classified by gender and by TV show preference (Sports, Documentary, News and Reality TV). [2]

The results are shown in the contingency table below.

	Sports	Documentary	News	Reality TV	Total
Male	20	24	32	11	87
Female	18	30	20	25	93
Total	38	54	52	36	180

Find the expected number of females who prefer documentary shows.

b. A χ^2 test at the $5\,\%$ significance level is used to determine whether TV show preference is independent of gender. [2]

Write down the p-value for the test.

c. State the conclusion of the test. Give a reason for your answer. [2]

Markscheme

a. $\frac{54}{180} \times \frac{93}{180} \times 180$ **OR** $\frac{54 \times 93}{180}$ ***(M1)***

$= 27.9$ ***(A1)*** ***(C2)***

b. $0.0321 \ (0.032139...)$ ***(A2)*** ***(C2)***

c. TV show preference is not independent of gender ***(A1)*****(ft)**

OR

reject the null hypothesis ***(A1)*****(ft)**

$0.0321 < 0.05$ ***(R1)*** ***(C2)***

Notes: Accept TV show preference is dependent on gender. Accept “associated”. Do not accept “correlated” or “related” or “linked”. Award ***(R1)*** for the comparison, ***(A1)*****(ft)** for a consistent conclusion with their answer to part (b). It is possible that **(A0)(R1)** be awarded. Do not award ***(A1)(R0)***.

Examiners report

a. Question 7: χ^2 test.

Candidates used their GDC to find the expected frequency with varying success whereas the p-value of the χ^2 test was usually correct; with some losing as many as four marks for giving answers to 1 significant figure with no working. As in the specimen paper the null hypotheses was not stated and so it was necessary to state what was being rejected. Candidates should write an explicit numerical comparison between p value and significance level to justify whether the null hypothesis is rejected or not. Amongst the candidates that made a comparison often the inequality sign was the wrong direction or the candidate made an inconsistent conclusion. There were many instances of poor mathematical terminology with correlation and independence used interchangeably likewise when candidates compared the significance level with their calculated χ^2 value.

b. Question 7: χ^2 test.

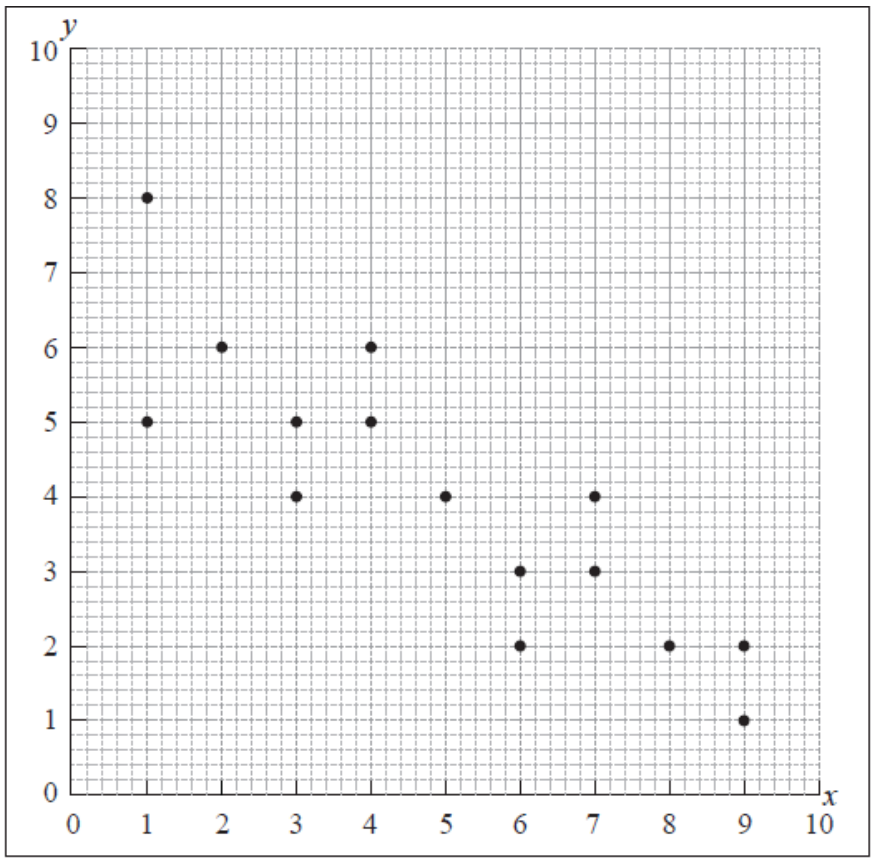
Candidates used their GDC to find the expected frequency with varying success whereas the p-value of the χ^2 test was usually correct; with some losing as many as four marks for giving answers to 1 significant figure with no working. As in the specimen paper the null hypotheses was not stated and so it was necessary to state what was being rejected. Candidates should write an explicit numerical comparison between p value and significance level to justify whether the null hypothesis is rejected or not. Amongst the candidates that made a comparison often the inequality sign was the wrong direction or the candidate made an inconsistent conclusion. There were many instances of poor mathematical terminology with correlation and independence used interchangeably likewise when candidates compared the significance level with their calculated χ^2 value.

c. Question 7: χ^2 test.

Candidates used their GDC to find the expected frequency with varying success whereas the p-value of the χ^2 test was usually correct; with some losing as many as four marks for giving answers to 1 significant figure with no working. As in the specimen paper the null hypotheses was not stated and so it was necessary to state what was being rejected. Candidates should write an explicit numerical comparison between p value and significance level to justify whether the null hypothesis is rejected or not. Amongst the candidates that made a comparison often the inequality sign was the wrong direction or the candidate made an inconsistent conclusion. There were many instances of poor mathematical terminology with correlation and independence used interchangeably likewise when candidates compared the significance level with their calculated χ^2 value.

Consider the following values of x and y and the scatter diagram which represents the information given in the table.

x	1	1	2	3	3	4	4	***b***	6	6	7	7	8	9	9
y	5	***a***	6	4	5	5	6	4	2	3	3	4	2	1	2



a. Write down the value of [2]

(i) a ;

(ii) b .

b. The mean of the x values is 5 and the mean of the y values is 4. Draw the line of best fit on the scatter diagram above. [2]

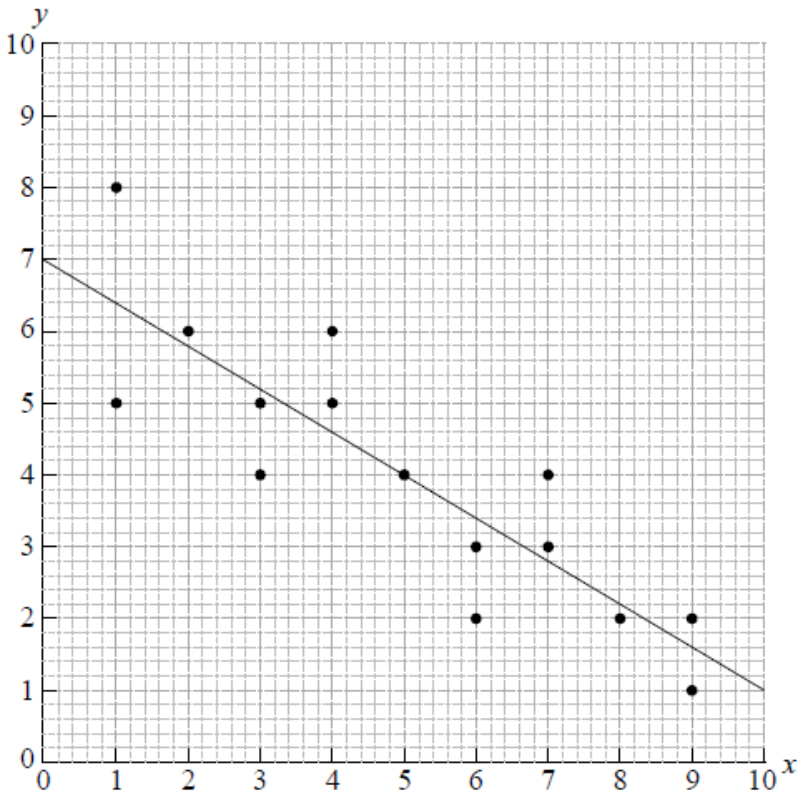
c. Use your line of best fit to estimate the value of y when $x = 6.5$. [2]

Markscheme

a. $a = 8$, $b = 5$ ***(A1)(A1)*** ***(C2)***

Note: Award ***(A0)(A1)*****(ft)** if $a = 5$, $b = 8$.

[2 marks]

b. 

(A1)(A1) ***(C2)***

Note: Award ***(A1)*** for straight line passing through (5, 4), ***(A1)*** for y-intercept between 6.8 and 8, if necessary with the line extended.

[2 marks]

c. 3.1 (±0.3) ***(M1)(A1)*****(ft)** ***(C2)***

Note: Award ***(M1)*** for an indication of use of their line of best fit (dotted lines or some indication of mark in the correct place on graph). Accept $y = 3.1 \pm 0.3$.

[2 marks]

Examiners report

a. Most candidates could find the values of a and b, draw the line of best fit and find an estimate of the value of y. Follow through marks could be awarded in part (c) with an incorrect line of best fit, if dotted lines were shown on their graphs, indicating their method.

b. Most candidates could find the values of a and b, draw the line of best fit and find an estimate of the value of y. Follow through marks could be awarded in part (c) with an incorrect line of best fit, if dotted lines were shown on their graphs, indicating their method.

c. Most candidates could find the values of a and b, draw the line of best fit and find an estimate of the value of y. Follow through marks could be awarded in part (c) with an incorrect line of best fit, if dotted lines were shown on their graphs, indicating their method.

The local park is used for walking dogs. The sizes of the dogs are observed at different times of the day. The table below shows the numbers of dogs present, classified by size, at three different times last Sunday.

	Small	Medium	Large
Morning	9	18	21
Afternoon	11	6	13
Evening	7	8	9

a. Write a suitable null hypothesis for a χ^2 test on this data. [1]

b. Write down the value of χ^2 for this data. [2]

c. The number of degrees of freedom is 4. Show how this value is calculated. [1]

d. The critical value, at the 5% level of significance, is 9.488. [2]

What conclusion can be drawn from this test? Give a reason for your answer.

Markscheme

a. H_0: The size of dog is independent of the time of day, (or equivalent) ***(A1)***

*Award **(A0)** for 'no correlation'* ***(C1)***

[1 mark]

b. $\chi^2 = 4.33$. *(accept 4.328)* ***(M1)(A1)***

*Note: GDC use is anticipated but candidates might calculate this by hand. **(M1)** can be awarded for a reasonable attempt to use the formula.* ***(C2)***

[2 marks]

c. $(3-1)(3-1) = 4$ ***(A1)***

Award mark for left hand side seen. ***(C1)***

[1 mark]

d. The hypothesis should not be rejected, *(allow 'accept H_0')*

OR

The size of dog is independent of the time of day ***(A1)*****(ft)**

$4.33 < 9.488$ *or* $0.363 > 0.05$ ***(R1)*****(ft)**

Allow $\chi_calc^2 < \chi_crit^2$ only if a value for χ_calc^2 is seen somewhere.

*Note: Award **(R1)(ft)** for comparing the values and **(A1)(ft)** if the conclusion is valid according to the comparison given. If no reason is given, or if the reason is wrong **both** marks are lost. Note that **(A0)(R1)(ft)** can be awarded but **(A1)(R0)** cannot.* ***(C2)***

[2 marks]

Examiners report

a. a) Most of the students got the null hypothesis right but quite a few used the word *correlation* instead of *independent*.

b) Candidates who used a GDC got it correct, while those who tried valiantly to calculate it by hand generally got an M1 but A0.

c) Most of the students knew how to calculate the degrees of freedom.

d) Many students did not have a clear idea which values to compare in order to arrive at a conclusion for the chi-squared test. Many compared the significance level with either the chi-squared value or the critical value. Some did not reject the hypothesis but either gave no reason or the wrong one.

As mentioned above, quite a number of candidates did not appear to have been taught this part of the course. There were many non-attempts. It was not a difficult question as indicated by the large number of candidates who scored full marks.

b. a) Most of the students got the null hypothesis right but quite a few used the word *correlation* instead of *independent.*

b) Candidates who used a GDC got it correct, while those who tried valiantly to calculate it by hand generally got an M1 but A0.

c) Most of the students knew how to calculate the degrees of freedom.

d) Many students did not have a clear idea which values to compare in order to arrive at a conclusion for the chi-squared test. Many compared the significance level with either the chi-squared value or the critical value. Some did not reject the hypothesis but either gave no reason or the wrong one.

As mentioned above, quite a number of candidates did not appear to have been taught this part of the course. There were many non-attempts. It was not a difficult question as indicated by the large number of candidates who scored full marks.

c. a) Most of the students got the null hypothesis right but quite a few used the word *correlation* instead of *independent.*

b) Candidates who used a GDC got it correct, while those who tried valiantly to calculate it by hand generally got an M1 but A0.

c) Most of the students knew how to calculate the degrees of freedom.

d) Many students did not have a clear idea which values to compare in order to arrive at a conclusion for the chi-squared test. Many compared the significance level with either the chi-squared value or the critical value. Some did not reject the hypothesis but either gave no reason or the wrong one.

As mentioned above, quite a number of candidates did not appear to have been taught this part of the course. There were many non-attempts. It was not a difficult question as indicated by the large number of candidates who scored full marks.

d. a) Most of the students got the null hypothesis right but quite a few used the word *correlation* instead of *independent.*

b) Candidates who used a GDC got it correct, while those who tried valiantly to calculate it by hand generally got an M1 but A0.

c) Most of the students knew how to calculate the degrees of freedom.

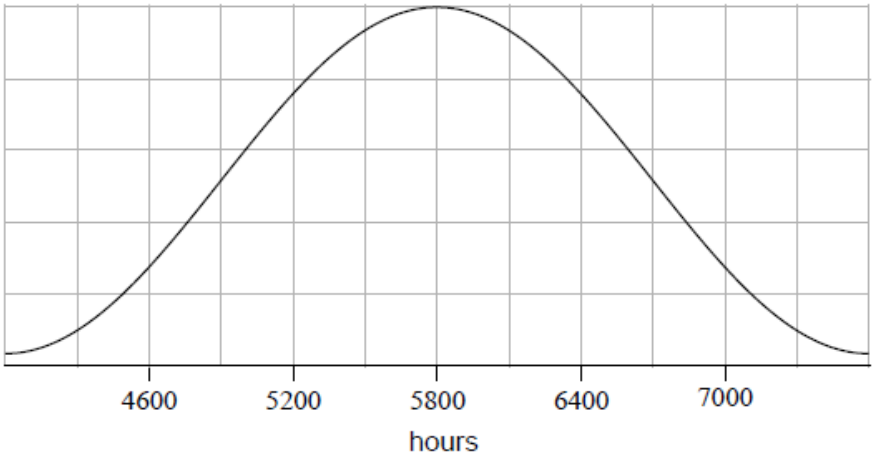
d) Many students did not have a clear idea which values to compare in order to arrive at a conclusion for the chi-squared test. Many compared the significance level with either the chi-squared value or the critical value. Some did not reject the hypothesis but either gave no reason or the wrong one.

As mentioned above, quite a number of candidates did not appear to have been taught this part of the course. There were many non-attempts. It was not a difficult question as indicated by the large number of candidates who scored full marks.

---

a. The lifetime, L , of light bulbs made by a company follows a normal distribution. [1]

L is measured in hours. The normal distribution curve of L is shown below.



Write down the mean lifetime of the light bulbs.

b. The standard deviation of the lifetime of the light bulbs is 850 hours. [2]

Find the probability that $5000 \leqslant L \leqslant 6000$, for a randomly chosen light bulb.

c. The company states that 90% of the light bulbs have a lifetime of at least k hours. [3]

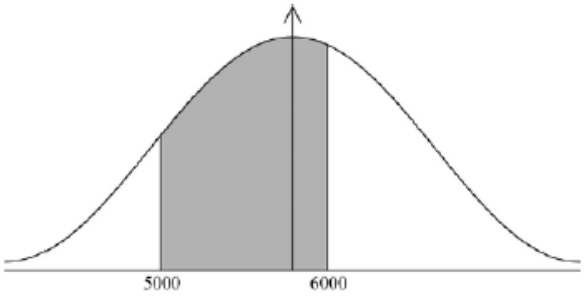
Find the value of k . Give your answer correct to the nearest hundred.

Markscheme

a. 5800 ***(A1)*** ***(C1)***

b. $0.420 \ (0.419703...)$ ***(A2)*(ft)** ***(C2)***

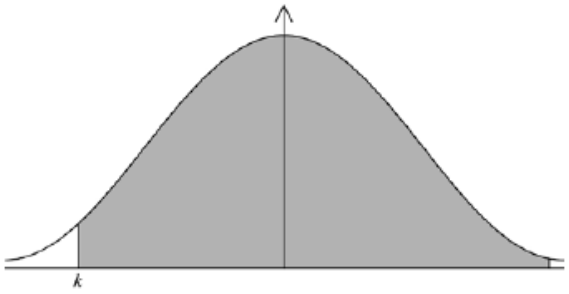
Note: Award ***(M1)*** for correct shaded area shown. Follow through from part (a).



c. $4710.68...$ ***(A2)*(ft)**

$= 4700$ ***(A1)*(ft)** ***(C3)***

Notes: Award ***(M1)*** for correct shaded area shown. Follow through from part (a). The final ***(A1)*(ft)** is awarded for the correct rounding of their answer to the nearest hundred hours.



Examiners report

a. Question 8: Normal distribution

Part (a) was correctly answered by the majority.

b. Part (b) was generally well attempted by those who had studied this part of the course. However, it was clear that many centres simply do not teach this part of the course.

c. In part (c), the two common faults were calculation of the bottom “tail” and incorrect rounding.

The weight, W, of bags of rice follows a normal distribution with mean 1000 g and standard deviation 4 g.

a. Find the probability that a bag of rice chosen at random weighs between 990 g and 1004 g. [2]

b. 95% of the bags of rice weigh less than k grams. [2]

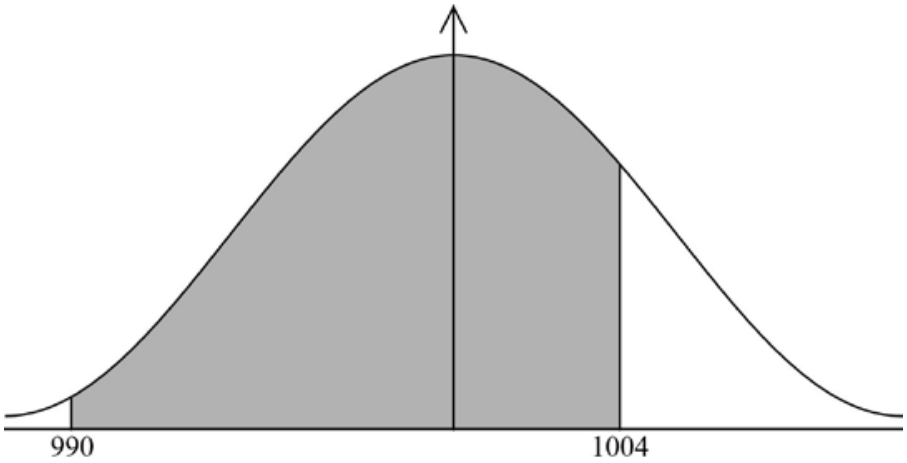
Find the value of k.

c. For a bag of rice chosen at random, $\mathrm{P}(1000 - a < W < 1000 + a) = 0.9$. [2]

Find the value of a.

Markscheme

a.

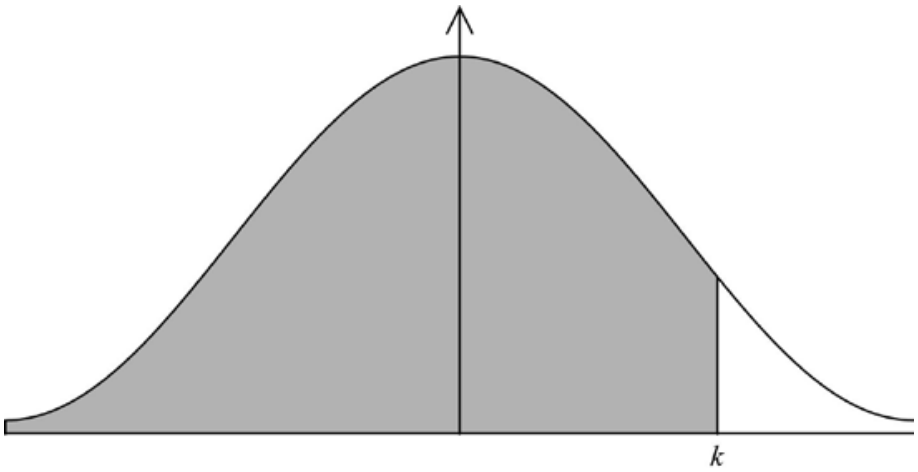


(M1)

Note: Award ***(M1)*** for approximate curve with 990 and 1004 in correct place.

$0.835 \quad (0.835135\ldots,\ 83.5\%)$ ***(A1)*** ***(C2)***

b.



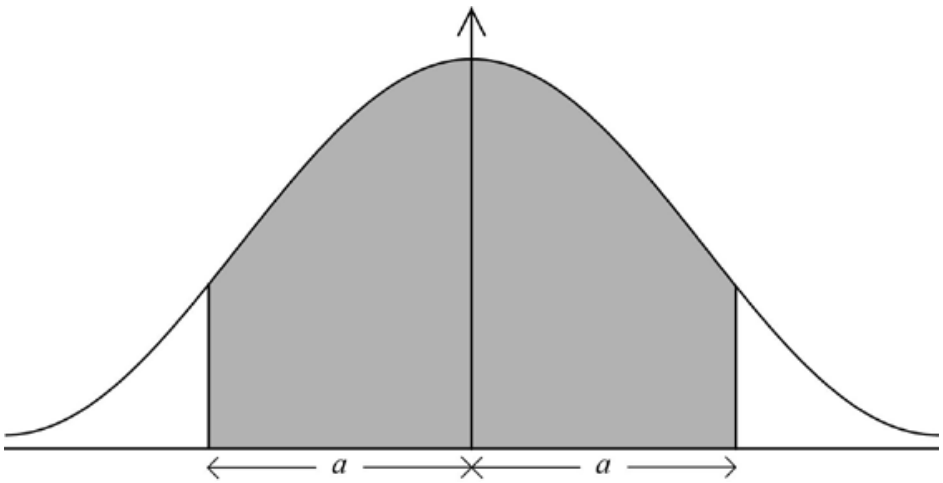
(M1)

Note: Award ***(M1)*** for approximate curve with k placed to the right of the mean.

$1010 \quad (1006.57\ldots)$ ***(A1)*** ***(C2)***

Note: Award full marks **only** for 1010, 1007 or an answer with more than 4 sf resulting from **correct** rounding of $1006.57\ldots$.

c.



(M1)

Note: Award ***(M1)*** for some indication of symmetry on diagram.

OR

$\mathrm{P}(W < 1000 - a) = 0.05$ **OR** $\mathrm{P}(W > 1000 + a) = 0.05$ ***(M1)***

Note: Award ***(M1)*** for probability with single inequality resulting from symmetry of diagram.

$(a =) \ 6.58 \quad (6.57941\ldots)$ ***(A1)*** ***(C2)***

Examiners report

a. A significant number of candidates did not answer this question. It was very rare that a correct method was shown for any of the parts of this question. Often a normal distribution graph was drawn with indication of the mean and multiples of the standard deviation, with indication of the corresponding probabilities, but not a diagram identifying the area under the curve corresponding to the questions. There were however many correct answers for part (a). For part (b) many answered incorrectly; the most common incorrect answer was 1008, resulting from adding 2 sd to the mean. Very few correct answers were given for part (c).

b. A significant number of candidates did not answer this question. It was very rare that a correct method was shown for any of the parts of this question. Often a normal distribution graph was drawn with indication of the mean and multiples of the standard deviation, with indication of the corresponding probabilities, but not a diagram identifying the area under the curve corresponding to the questions. There were however many correct answers for part (a). For part (b) many answered incorrectly; the most common incorrect answer was 1008, resulting from adding 2 sd to the mean. Very few correct answers were given for part (c).

c. A significant number of candidates did not answer this question. It was very rare that a correct method was shown for any of the parts of this question. Often a normal distribution graph was drawn with indication of the mean and multiples of the standard deviation, with indication of the corresponding probabilities, but not a diagram identifying the area under the curve corresponding to the questions. There were however many correct answers for part (a). For part (b) many answered incorrectly; the most common incorrect answer was 1008, resulting from adding 2 sd to the mean. Very few correct answers were given for part (c).

Malthouse school opens at 08:00 every morning.

The daily arrival times of the 500 students at Malthouse school follow a normal distribution. The mean arrival time is 52 minutes after the school opens and the standard deviation is 5 minutes.

a.i. Find the probability that a student, chosen at random arrives at least 60 minutes after the school opens. [2]

a.ii. Find the probability that a student, chosen at random arrives between 45 minutes and 55 minutes after the school opens. [2]

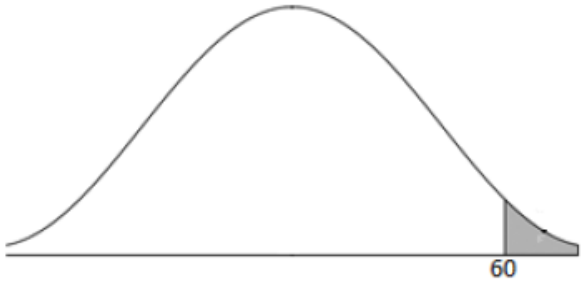
b. A second school, Mulberry Park, also opens at 08:00 every morning. The arrival times of the students at this school follows exactly the same distribution as Malthouse school. [2]

Given that, on one morning, 15 students arrive at least 60 minutes after the school opens, estimate the number of students at Mulberry Park school.

Markscheme

a.i. 0.0548 (0.054799…, 5.48%) ***(A2) (C2)***

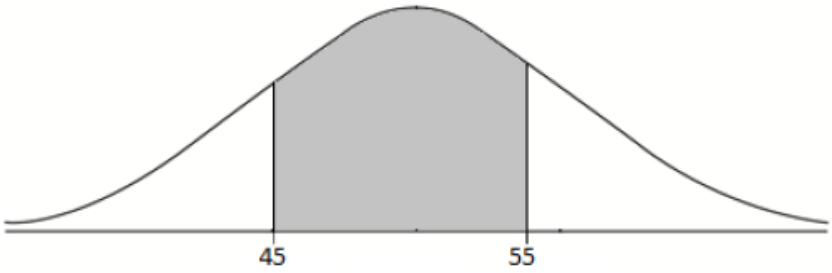
Note: Award ***(M1)(A0)*** for a correct probability statement, $P(X > 60)$, or normal distribution graph with correctly shaded region, leading to incorrect or no answer.



[2 marks]

a.ii. 0.645 (0.6449900…, 64.5%) ***(A2) (C2)***

Note: Award ***(M1)(A0)*** for a correct probability statement, $P(45 < X < 55)$, or normal distribution graph with correctly shaded region, leading to incorrect or no answer.



[2 marks]

b. $\frac{15}{0.0548}$ **(M1)**

Note: Award ***(M1)*** for dividing 15 by their part (a)(i).

Accept an equation of the form $15 = x \times 0.0548$ for ***(M1)***.

274 (273.722…) ***(A1)*(ft) *(C2)***

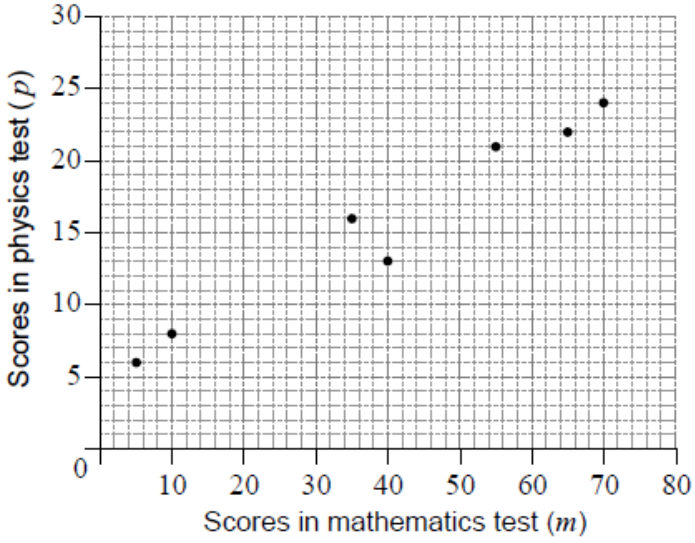
Note: Follow through from part (a)(i). Accept 273.

[2 marks]

Examiners report

a.i. [N/A]
a.ii. [N/A]
b. [N/A]

The following scatter diagram shows the scores obtained by seven students in their mathematics test, *m*, and their physics test, *p*.



The mean point, M, for these data is (40, 16).

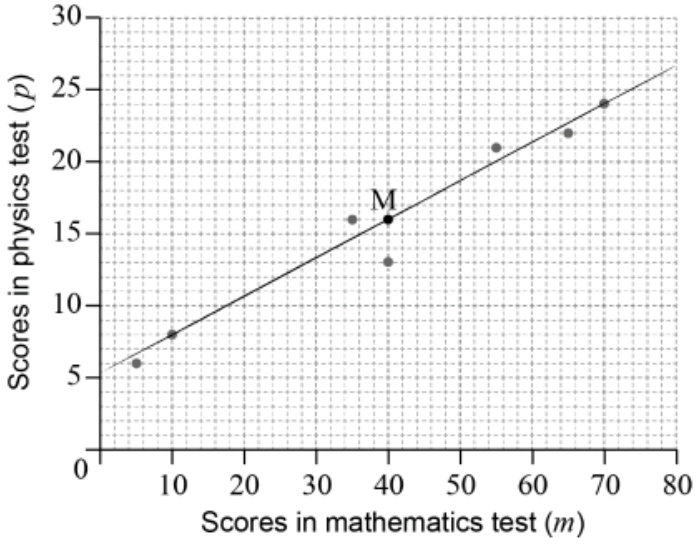
a. Plot and label the point M$(\bar{m},\ \bar{p})$ on the scatter diagram. [2]

b. Draw the line of best fit, by eye, on the scatter diagram. [2]

c. Using your line of best fit, estimate the physics test score for a student with a score of 20 in their mathematics test. [2]

Markscheme

a.



(A1)(A1) (C2)

Note: Award ***(A1)*** for mean point plotted and ***(A1)*** for labelled M.

[2 marks]

b. straight line through their mean point crossing the p-axis at 5±2 ***(A1)(ft)(A1)(ft) (C2)***

Note: Award ***(A1)*****(ft)** for a straight line through their mean point. Award ***(A1)*****(ft)** for a correct p-intercept if line is extended.

[2 marks]

c. point on line where m = 20 identified and an attempt to identify y-coordinate ***(M1)***

10.5 ***(A1)*****(ft)** ***(C2)***

Note: Follow through from their line in part (b).

[2 marks]

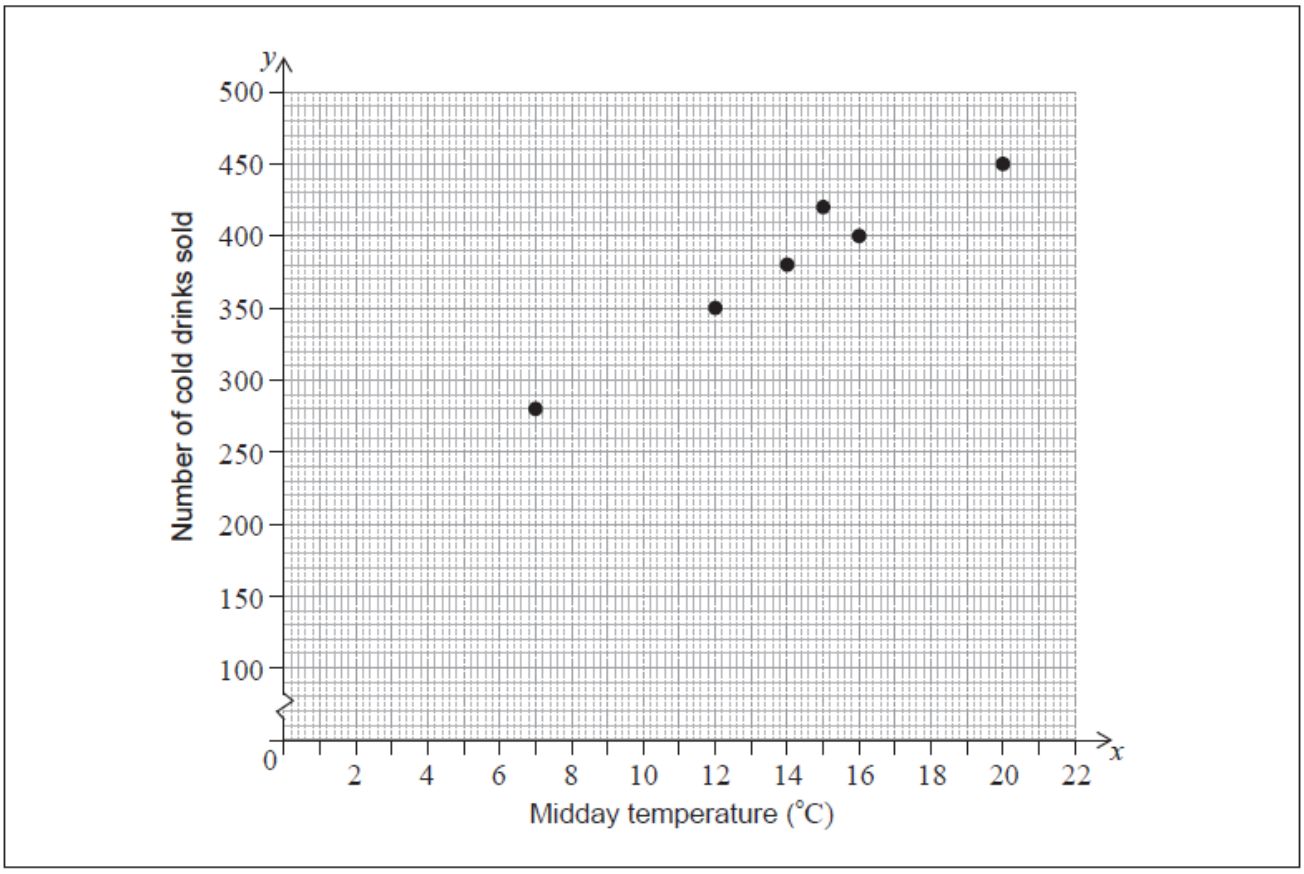
Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

---

a. Each day a supermarket records the midday temperature and how many cold drinks are sold on that day. The following table shows the supermarket's data for the last 6 days. This data is also shown on a scatter diagram. [2]

Midday temperature, °C (x)	7	12	14	15	16	20
Number of cold drinks sold (y)	280	350	380	420	400	450



Write down

i) the mean temperature, $\bar{x}$;

ii) the mean number of cold drinks sold, $\bar{y}$.

b. Draw the line of best fit on the scatter diagram. [2]

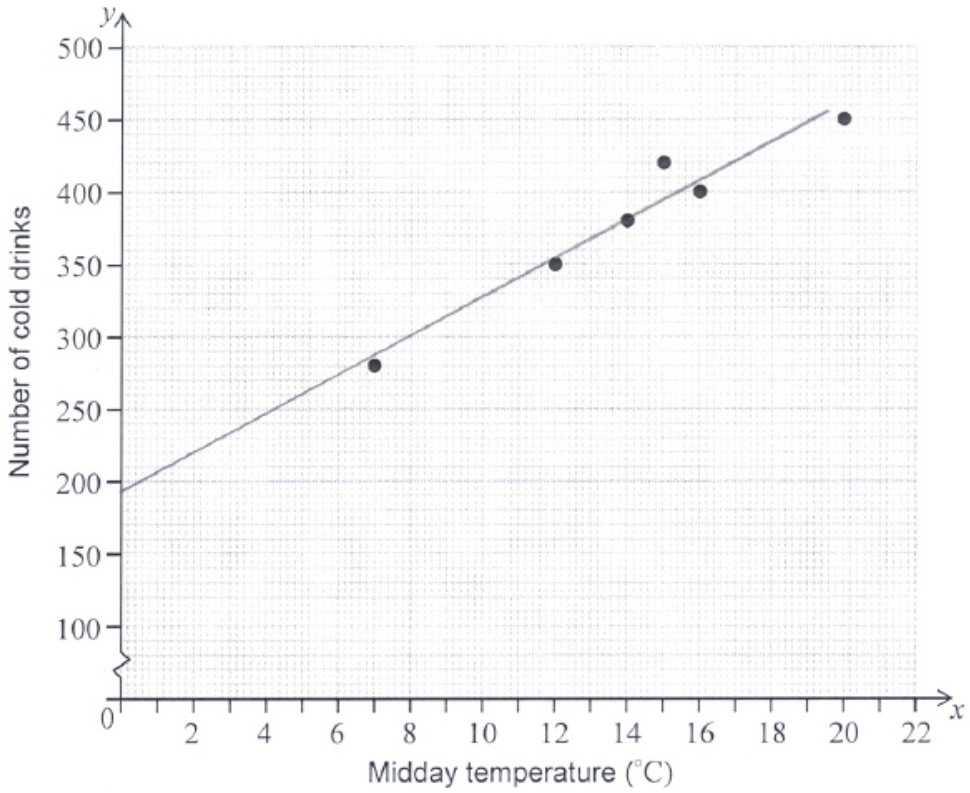
c. Use the line of best fit to estimate the number of cold drinks that are sold on a day when the midday temperature is $10\,^{\circ}\mathrm{C}$. [2]

Markscheme

a. i) 14 ***(A1)***

ii) 380 ***(A1)*** ***(C2)***

b.



(A1)**(ft)*****(A1)*** ***(C2)***

Note: Award ***(A1)*****(ft)** for a **straight** line going through **their** mean point, ***(A1)*** for intercepting the y-axis between 160 and 220 inclusive. Follow through from part (a).

c. an attempt to use their line of best fit to find y value at $x = 10$ ***(M1)***

Note: Award ***(M1)*** for an indication of use of their line of best fit (dotted lines or some indication of mark in the correct place on graph).

OR

$13.4\,(10) + 192$ ***(M1)***

Note: Award ***(M1)*** for correct substitution into the regression equation, $y = 13.4x + 192$.

$= 326$ ***(A1)*****(ft)** ***(C2)***

Note: Follow through from part (b). Accept answers between 310 and 340, inclusive.

Examiners report

a. Question 9: Linear regression.

The correct means were usually written down. Many candidates drew a line of best fit that did not go through their $(\bar{x},\ \bar{y})$. Almost all candidates were able to use the line of best fit (either the one they had drawn or the regression line found using their GDC) to make a reasonable estimate. Feedback from teachers suggests that many are using line of best fit and line of regression as synonyms. This is not the case; both are explicitly mentioned in the guide and candidates are expected to understand both terms.

b. Question 9: Linear regression.

The correct means were usually written down. Many candidates drew a line of best fit that did not go through their $(\bar{x},\ \bar{y})$. Almost all candidates were able to use the line of best fit (either the one they had drawn or the regression line found using their GDC) to make a reasonable estimate. Feedback from teachers suggests that many are using line of best fit and line of regression as synonyms. This is not the case; both are explicitly mentioned in the guide and candidates are expected to understand both terms.

c. Question 9: Linear regression.

The correct means were usually written down. Many candidates drew a line of best fit that did not go through their $(\bar{x},\ \bar{y})$. Almost all candidates were able to use the line of best fit (either the one they had drawn or the regression line found using their GDC) to make a reasonable estimate. Feedback from teachers suggests that many are using line of best fit and line of regression as synonyms. This is not the case; both are explicitly mentioned in the guide and candidates are expected to understand both terms.