

Please check the examination details below before entering your candidate information

Candidate surname					Other names				
Centre Number					Candidate Number				
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	

Pearson Edexcel Level 3 GCE

Paper reference	8MA0/21
-----------------	----------------

Mathematics

Advanced Subsidiary

PAPER 21: Statistics

<p>You must have: Mathematical Formulae and Statistical Tables (Green), calculator</p>	<p>Total Marks</p>
---	--------------------

Candidates may use any calculator permitted by Pearson regulations. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulae stored in them.

Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B).
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided
– *there may be more space than you need.*
- You should show sufficient working to make your methods clear.
Answers without working may not gain full credit.
- Values from statistical tables should be quoted in full. If a calculator is used instead of tables the value should be given to an equivalent degree of accuracy.
- Inexact answers should be given to three significant figures unless otherwise stated.

Information

- A booklet 'Mathematical Formulae and Statistical Tables' is provided.
- The total mark for this part of the examination is 30. There are 5 questions.
- The marks for **each** question are shown in brackets
– *use this as a guide as to how much time to spend on each question.*

Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

Turn over ►

P69599A

©2022 Pearson Education Ltd.

Q:1/1/1/



1. The relationship between two variables p and t is modelled by the regression line with equation

$$p = 22 - 1.1t$$

The model is based on observations of the independent variable, t , between 1 and 10

- (a) Describe the correlation between p and t implied by this model.

(1)

Given that p is measured in centimetres and t is measured in days,

- (b) state the units of the gradient of the regression line.

(1)

Using the model,

- (c) calculate the change in p over a 3-day period.

(2)

Tisam uses this model to estimate the value of p when $t = 19$

- (d) Comment, giving a reason, on the reliability of this estimate.

(1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 1 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 1 is 5 marks)

2. A manufacturer of sweets knows that 8% of the bags of sugar delivered from supplier *A* will be damp.

A random sample of 35 bags of sugar is taken from supplier *A*.

- (a) Using a suitable model, find the probability that the number of bags of sugar that are damp is

(i) exactly 2

(ii) more than 3

(3)

Supplier *B* claims that when it supplies bags of sugar, the proportion of bags that are damp is less than 8%

The manufacturer takes a random sample of 70 bags of sugar from supplier *B* and finds that only 2 of the bags are damp.

- (b) Carry out a suitable test to assess supplier *B*'s claim.

You should state your hypotheses clearly and use a 10% level of significance.

(4)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 2 continued

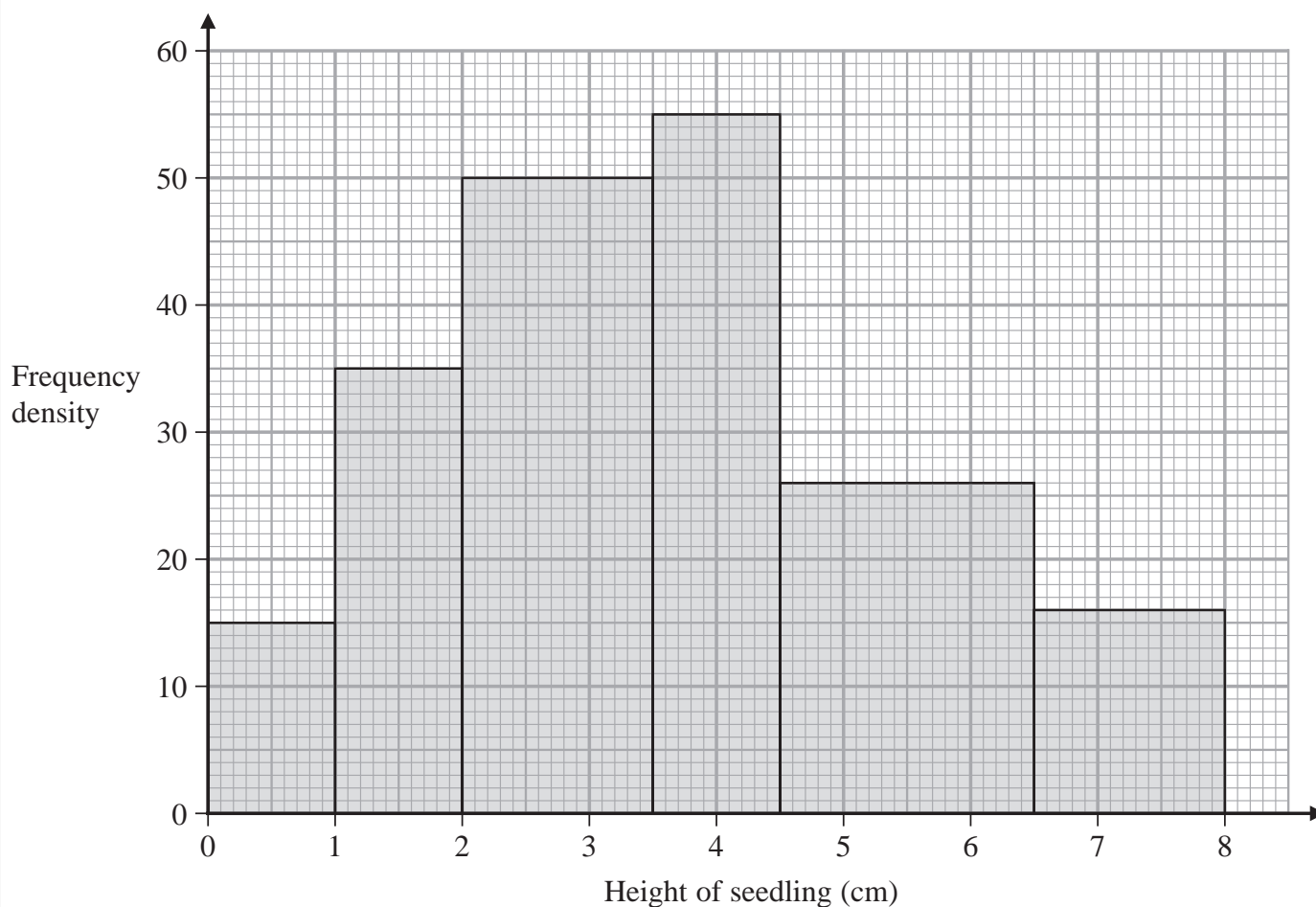
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 2 is 7 marks)

3. The histogram summarises the heights of 256 seedlings two weeks after they were planted.



DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

- (a) Use linear interpolation to estimate the median height of the seedlings.

(4)

Chris decides to model the **frequency density** for these 256 seedlings by a curve with equation

$$y = kx(8 - x) \quad 0 \leq x \leq 8$$

where k is a constant.

- (b) Find the value of k

(3)

Using this model,

- (c) write down the median height of the seedlings.

(1)



Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Question 3 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

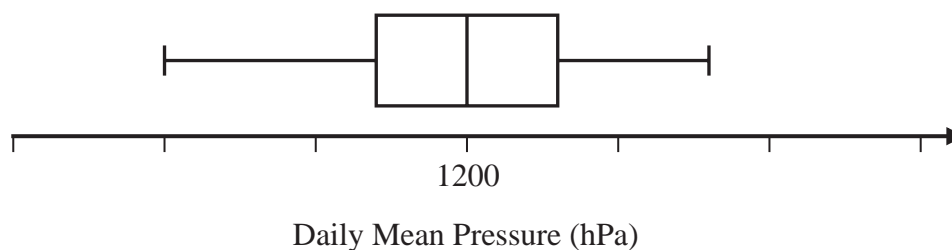
DO NOT WRITE IN THIS AREA

(Total for Question 3 is 8 marks)

4. Jiang is studying the variable Daily Mean Pressure from the large data set.

He drew the following box and whisker plot for these data for one of the months for one location using a linear scale but

- he failed to label all the values on the scale
- he gave an incorrect value for the median



Using your knowledge of the large data set, suggest a suitable value for

- (a) the median, (1)
- (b) the range. (1)

(You are not expected to have memorised values from the large data set. The question is simply looking for sensible answers.)

(Total for Question 4 is 2 marks)



- The probability distributions for R and G are given below.

g	1	4
$P(G = g)$	$\frac{2}{3}$	$\frac{1}{3}$

(3)

$$P(X = 20) = \frac{1}{6} \quad \text{and} \quad P(X = 50) = \frac{1}{4}$$

- (5)

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



P 6 9 5 9 9 A 0 1 1 1 2

Question 5 continued

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

(Total for Question 5 is 8 marks)**TOTAL FOR STATISTICS IS 30 MARKS**



Mark Scheme (Results)

Summer 2022

Pearson Edexcel GCE Mathematics

Advanced Subsidiary Level in Mathematics

Paper 21 8MA0/21 Statistics

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

June 2022

Publications Code 8MA0_21_2206_MS

All the material in this publication is copyright

© Pearson Education Ltd 2022

General Marking Guidance

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied positively. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is no ceiling on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL GCE MATHEMATICS

General Instructions for Marking

1. The total number of marks for the paper is 75.
2. The Edexcel Mathematics mark schemes use the following types of marks:
 - **M marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.**
 - **A marks:** Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
 - **B marks** are unconditional accuracy marks (independent of M marks)
 - Marks should not be subdivided.
3. Abbreviations

These are some of the traditional marking abbreviations that will appear in the mark schemes.

- bod – benefit of doubt
 - ft – follow through
 - the symbol \surd will be used for correct ft
 - cao – correct answer only
 - cso – correct solution only. There must be no errors in this part of the question to obtain this mark
 - isw – ignore subsequent working
 - awrt – answers which round to
 - SC: special case
 - oe – or equivalent (and appropriate)
 - dep – dependent
 - indep – independent
 - dp decimal places
 - sf significant figures
 - * The answer is printed on the paper
 - \square The second mark is dependent on gaining the first mark
4. **All A marks are 'correct answer only' (cao.),** unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

Qu	Scheme	Marks	AO
1. (a)	Negative (since gradient of regression line is negative)	B1 (1)	1.2
(b)	cm/day (o.e. e.g. cm day^{-1})	B1 (1)	2.2a
(c)	$3 \times [\pm] 1.1$ = decrease of 3.3 [cm]	M1 A1 (2)	3.4 1.1b
(d)	19 is (well) outside the range [1, 10] <u>or</u> involves extrapolation (o.e.) so (possibly) unreliable/ inaccurate (o.e.)	B1 (1) (5 marks)	2.4
Notes			
Answers may be written within the question.			
(a)	B1 for stating “negative”. Allow a correct interpretation e.g. as t increases then p decreases (o.e.) [ignore any values] B0 for contradictory statements e.g. “negative correlation since as t increases p increases”		
(b)	B1 for a correct description of the units (allow fraction, /, or “per” and allow “d” for “day”)		
(c)	M1 for attempt at a calculation (allow use of $t = x$ and $t = x + 3$ followed by subtraction that should lead to 3.3) A1 for correct description must include word “decrease” (o.e.) and value “3.3” Just seeing: $22 - 1.1 \times 3 = 18.7$ is M0A0 BUT going on to subtract 18.7 from 22 scores M1 Reaching 3.3 <u>and</u> stating “decrease” or “reduced” (o.e.) will score the A1 too An answer of -3.3 without a word describing decrease (o.e.) will just score M1A0		
(d)	B1 for stating “unreliable” (o.e.) and giving a suitable reason based on idea of extrapolation Must have both statement about reliability and suitable reason e.g. $t = 19$ is too big <u>or</u> (Model is based on) t between 1 and 10 (only) [since this implies $t = 19$ is too big] Allow e.g. (model) “may not work” because of “extrapolation” Just saying “no” since “extrapolation” is B0 but “unreliable”(o.e.) since “extrapolation” is B1		

Qu	Scheme	Mark	AO
2. (a)	$[D = \text{number of bags that are damp}] \quad D \sim B(35, 0.08) \quad \text{NB } 0.08 = \frac{2}{25}$	M1	3.3
(i)	$P(D = 2) = 0.2430497... \quad \text{awrt } \underline{\mathbf{0.243}}$	A1	3.4
(ii)	$P(D > 3) = [1 - P(D \leq 3) = 1 - 0.69397...] = 0.30602... \quad \text{awrt } \underline{\mathbf{0.306}}$	A1	1.1b
		(3)	
(b)	$H_0 : p = 0.08 \quad H_1 : p < 0.08$	B1	2.5
	$[X \sim] B(70, 0.08)$	M1	2.1
	$[P(X \leq 2)] = 0.0739756... \quad \text{awrt } \underline{\mathbf{0.074}}$	A1	1.1b
	$[0.074 < 0.10 \text{ so significant, reject } H_0 \text{ so...}]$		
	there <u>is</u> evidence to <u>support</u> supplier <u>B's claim</u> (o.e.)	A1	2.2b
		(4)	
		(7 marks)	
Notes			
(a)	M1 for selecting a correct model: sight of or use of $B(35, 0.08)$ [Condone $B(0.08, 35)$] May be implied by one correct answer or sight of $P(D \leq 3) = \text{awrt } 0.694$ (or allow 0.693) or seeing $\binom{35}{2} 0.08^2 \times (1 - 0.08)^{35-2}$ Saying $B(35, 8\%)$ without a correct calculation would score M0		
(i)	1 st A1 for awrt 0.243		
(ii)	2 nd A1 for awrt 0.306 (Condone poor use of notation e.g. $P(D = 3) = 0.306...$ i.e. just mark ans)		
NB	$P(D \leq 3) = 0.539$ scores 2 nd A0 but would of course score M1		
(b)	B1 for both hypotheses correct in terms of p or π [Condone 8% for 0.08] M1 for sight or correct use of $B(70, 0.08)$ [Condone $B(0.08, 70)$] May be implied by prob of 0.074 or better 1 st A1 for final answer awrt 0.074 can condone poor notation e.g. $P(X = 2) = \text{awrt } 0.074$ Can allow this mark for CR of $X \leq 2$ provided $[P(X \leq 2)] = 0.074$ (or better) is seen [Can allow 0.07 if $X \sim B(70, 0.08)$ and $P(X \leq 2)$ are both seen] 2 nd A1 (dep on M1A1 but independent of hypotheses) for a correct inference in context Must mention <u>claim</u> or <u>B</u> and idea of <u>support for</u> ... or <u>proportion/probability</u> (of damp bags) and idea of <u>less</u> than 8% or A 2 nd A0 for contradictory statements e.g. "accept H_0 so evidence to support B's claim" 2 nd A0 if you see $0.0739... < 0.08$ so significant/ reject H_0 etc		
MR	0.8 for 0.08 In (a) allow M1 for $B(35, 0.8)$ then A0A0 In (b) allow B1 for Hypotheses and M1 for $B(70, 0.8)$ seen, then A0A0		

Qu	Scheme	Mark	AO															
3. (a)	<table><tr><th>Class</th><th>Frequency</th><th>Cum. Frequency</th></tr><tr><td>0 – 1</td><td>15</td><td>15</td></tr><tr><td>1 – 2</td><td>35</td><td>50</td></tr><tr><td>2 – 3.5</td><td>75</td><td>125</td></tr><tr><td>3.5 – 4.5</td><td>55</td><td>180</td></tr></table>	Class	Frequency	Cum. Frequency	0 – 1	15	15	1 – 2	35	50	2 – 3.5	75	125	3.5 – 4.5	55	180	M1	2.1
	Class	Frequency	Cum. Frequency															
	0 – 1	15	15															
	1 – 2	35	50															
	2 – 3.5	75	125															
	3.5 – 4.5	55	180															
		A1	1.1b															
	$[Q_2 =](3.5) + \frac{\frac{256}{2} - "125"}{"55"} \times (4.5 - 3.5)$ <u>or</u> $(4.5) - \frac{"180" - \frac{256}{2}}{"55"} \times 1$ $= 3.5545\ldots$ awrt <u>3.55</u>	M1	2.1															
		A1	1.1b															
		(4)																
(b)	Need area under curve to be 256 so $\int_{(0)}^{(8)} kx(8 - x) \, dx = 256$	M1	3.1a															
	$k \left[4x^2 - \frac{x^3}{3} \right]_{(0)}^{(8)} = 256$	M1	1.1b															
	$\left\{ k \left[4 \times 8^2 - \frac{8}{3} \times 8^2 \right] = 256 \Rightarrow \right\}$ <u>k = 3</u>	A1	1.1b															
		(3)																
(c)	[By symmetry median =] <u>4</u>	B1	2.2a															
		(1)																
		(8 marks)																
Notes																		
(a)	1 st M1 for an attempt to form frequency table (at least 1 st 4 rows and freq <u>or</u> cum freq seen must have the frequency of 75 correct and can condone one error/omission in 15, 35, 55) Frequencies or cum freq may be seen on bars of the histogram 1 st A1 for identifying class, freq and cum freq (i.e. highlighted values from the table) <u>or</u> sight of 3.5-4.5, freq of 55 and “128” – 125 or 180 – “128” <u>or</u> diagram with 125, “128”, 180, 3.5 & 4.5 May be implied by values in 2 nd M1 expression 2 nd M1 for a correct calculation for Q_2 (condone error in end point e.g. 3.45 or 3.49 etc) Can fit their “125” (provided > 100) and their “55” Allow use of $(n + 1)$, usually see 128.5 – ... leading to 3.5636... or awrt 3.56 2 nd A1 awrt 3.55 but 3.555 is fine (allow 3.56 if $(n + 1)$ being used ...need sight of $\frac{257}{2}$ etc) Correct answer with no incorrect working scores 4/4																	
(b)	1 st M1 for identifying the need to find the area under the curve by integrating 2 nd M1 for correct integration and = 256 (condone missing limits) A1 for $k = 3$ [May see use of calculator for the integration so score 2 nd M1A1 together]																	
(c)	NB The answer to part (c) may be written within the question. B1 for 4 (Independent of their value of k but must be their “ x ” value) NB when $k = 0.25$ and $x = 4$ gives $y = 4$ so must be clear they intend median = 4 The statement in part (c) “ $k = 4$ ” is B0																	

	Scheme	Marks	AO
--	--------	-------	----

4. (a)	Accept 990 to 1030 inclusive	B1 (1)	1.1b
(b)	Any range between 10 and 50 inclusive	B1 (1)	1.1b
		(2 marks)	
Notes			
(a)	B1 (Median pressures usually around 1000~1020)	[LDS mark]	
(b)	B1 Any answer in this range Allow answers in the form $a \sim b$ where $ b - a $ is between 10 and 50 Also allow the case where <u>both</u> a and b are in $[10, 50]$	[LDS mark]	

Qu	Scheme	Mark	AO
5. (a)(i)	Require $R = 3$ and $G = 4$ so probability is $\frac{3}{4} \times \frac{1}{3}$	M1	2.1
	$= \frac{1}{4}$ or <u>0.25</u>	A1	1.1b
	(ii)		
	$[R \text{ must be } 2 \text{ and } G = 1 \text{ so } \frac{1}{4} \times \frac{2}{3}] = \frac{1}{6}$	A1	1.1b
	(b)		
	$P(X = 50) = 0.25$ must mean $R = 3$ and $G = 4$	M1	3.1a
	so $3m + 4n = 50$	A1	1.1b
	$P(X = 20) = \frac{1}{6} \Rightarrow R = 2, G = 1$ so $2m + n = 20$	A1	2.1
	Solving: $3m + 4(20 - 2m) = 50$ (o.e.)	M1	1.1b
	<u>$m = 6$</u> and <u>$n = 8$</u>	A1	3.2a
		(5)	
		(8 marks)	
	Notes		
(a)(i)	M1 for sight of $\frac{3}{4} \times \frac{1}{3}$ or $\frac{1}{4} \times \frac{2}{3}$ as a single product BUT allow e.g. $\frac{3}{4} \times \frac{1}{3} + \frac{1}{3} \times \frac{3}{4}$ to score M1 However if the products are later added e.g. $\frac{3}{4} \times \frac{1}{3} + \frac{1}{4} \times \frac{2}{3}$ it is M0 May be implied by one correct answer to (i) or (ii)		
	A1 for $\frac{1}{4}$ or 0.25 or exact equivalent (allow 25%)		
(ii)	A1 for $\frac{1}{6}$ or exact equivalent		
(b)	For the 1st 4 marks condone incorrect labelling e.g. R for m or G for n if intention is clear 1 st M1 for identifying either set of cases ($R = 2, G = 1, X = 20$) or ($R = 3, G = 4, X = 50$) Allow 1 st M1 for $P(X = 20) = \frac{1}{4} \times \frac{2}{3}$ or $P(X = 50) = \frac{3}{4} \times \frac{1}{3}$ NOT just $P(X = 20) = \frac{1}{6}$ etc or $\frac{1}{4}m + \frac{2}{3}n = 20$ or $\frac{3}{4}m + \frac{1}{3}n = 50$ and might score 2 nd M1 (answer is $m = 64, n = 6$) or $\frac{1}{4}m + \frac{2}{3}n = \frac{1}{6}$ or $\frac{3}{4}m + \frac{1}{3}n = \frac{1}{4}$ and might score 2 nd M1 (answer is $m = \frac{4}{15}, n = \frac{3}{20}$) or $2m + n = \frac{1}{6}$ or $3m + 4n = \frac{1}{4}$ and might score 2 nd M1 (answer is $m = \frac{1}{12}, n = 0$) or $2m + n = 50$ and $3m + 4n = 20$ and might score 2 nd M1 (answer is $m = 36, n = -22$) 1 st A1 for one correct equation 2 nd A1 for both correct equations and no incorrect equations, unless they attempt to solve the correct 2 equations only 2 nd M1 for attempt to solve their two linear equations in m and n (reduce to an equation in one variable, condone one sign error). May be implied by $m = 6$ and $n = 8$. If they use one of the 4 sets of equations for 1 st M1 and use a calculator to write down the answer, we will allow this mark for sight of the correct answers to those equations as given above.		
Calc	3 rd A1 $m = 6$ and $n = 8$ only (no incorrect labelling here) Correct answer by trial can score 5/5 if no incorrect working seen.		