Class Adm No

Candidate Name: ______________________________

H

millennia institute

2017 Promotional Examination II
Pre-University 2

MATHEMATICS **8865/01**

Paper 1 **12 September 2017**

3 hours

Additional Materials: Answer Paper
List of Formulae (MF 26)

READ THESE INSTRUCTIONS FIRST

Write your name and class on all the work you hand in.
Write in dark blue or black pen on both sides of the paper.
You may use a soft pencil for any diagrams or graphs.
Do not use paper clips, highlighters, glue or correction fluid.

Answer **all** the questions.
Give non-exact numerical answers correct to 3 significant figures, or 1 decimal place in the case of angles in degrees, unless a different level of accuracy is specified in the question.
You are expected to use an approved graphing calculator.
Unsupported answers from a graphing calculator are allowed unless a question specifically states otherwise.
Where unsupported answers from a graphing calculator are not allowed in a question, you are required to present the mathematical steps using mathematical notations and not calculator commands.
You are reminded of the need for clear presentation in your answers.

At the end of the examination, arrange your answers in NUMERICAL ORDER and fasten all your work securely together.

The number of marks is given in brackets [ ] at the end of each question or part question.

This question paper consists of 6 printed pages.

Section A: Pure Mathematics [40 marks]

1 Find the range of values of k for which the equation $x^2+2kx+3k+4=0$ has real roots. [3]

2 Find $\dfrac{\mathrm{d}}{\mathrm{d}x}\left[\ln\left(\dfrac{\sqrt{2x^2+1}}{2-x}\right)\right]$. [3]

3 The graph of $y=ax^2+bx+c$ passes through the points (1, 6), (7, 234) and (13, 822). By forming a system of linear equations, find the values of a, b and c. [3]

Hence find the range of values of x for which the graph is decreasing. [2]

4 The curve C has equation $y=1-\mathrm{e}^{-2x}$.

(i) Sketch the graph of C, stating the equation(s) of any asymptote(s). [2]

(ii) Find the equation of the tangent to the curve C at $x=1$, giving your answer in the form $y=mx+c$, where m and c are exact constants to be found. [3]

(iii) Find the exact area bounded by the curve C, the line x = 1 and the x-axis. Deduce the exact area bounded by the curve C, y = $1-\mathrm{e}^{-2}$ and the y-axis. [3]

5 **(i)** Sketch, on the same diagram, the graphs of $y=\ln(x+3)$ and $y=\dfrac{2x^2-1}{x^2}$, labelling clearly the equations of any asymptotes. There is no need to find the coordinates of any points where the graphs cross the axes. [4]

(ii) Find the x-coordinate(s) of the point(s) of intersection of the graphs of $y=\ln(x+3)$ and $y=\dfrac{2x^2-1}{x^2}$. Hence solve the inequality $1+x^2\ln(x+3)<2x^2$. [3]

(iii) Find the area bounded by the two curves $y=\ln(x+3)$ and $y=\dfrac{2x^2-1}{x^2}$, giving your answer to 3 decimal places. [2]

6 The managing director of a company tracked the rate of output, x units per month, of its product regularly over t months. His analyst believes that x and t can be modelled by the equation

$$x = a + 30t^2 - 2t^3,$$

where $0 \le t \le 12$ and a is a positive constant.

(i) Using differentiation, find the maximum rate of output in the year in terms of a, justifying that this is a maximum. [4]

(ii) Sketch the graph of x against t for a = 25 and give an interpretation of the value of a. [3]

(iii) Find the exact area of the region bounded by the graph, the axes and the line t = 12 for a = 25. Give an interpretation of the value of this area. [3]

The analyst also believes that the profit per month, \$$y$ million, can be modelled by the equation

$$y = \ln[(x-a)^2 + (x-a)], \text{ where } x > a.$$

(iv) By expressing y in terms of t, find the rate at which the profit per month is increasing when t = 1. [2]

Section B: Probability and Statistics [60 marks]

7 Consider arranging all the letters of the word **FORMULAE**.

(i) Find the number of different arrangements if there are no restrictions. [1]

(ii) Find the probability that the arrangement starts and ends with a consonant and the vowels are together. [3]

Codewords are formed by arranging 3 letters from the letters of the word **FORMULAE**.

(iii) Find the number of different codewords that can be formed. [2]

8 In a game, there are three boxes A, B and C. Box A contains 1 red and 9 white balls. Box B contains 2 red and 8 white balls. Box C contains 3 red and 7 white balls. All the red and white balls in the three boxes are indistinguishable other than the colours.

The player selects one of the three boxes by tossing a fair die. The player selects a ball from Box A if the die shows 1, 2 or 3. The player selects a ball from Box B if the die shows 4 or 5. The player selects a ball from Box C if the die shows a 6.

If the player selects a white ball, the game is over. If the player selects a red ball, the person wins \$100 and is allowed to draw another ball from the same box containing the remaining balls. If the second ball is white, the game is over. If the second ball is red, the player wins another \$200 and is allowed to draw another ball from the same box containing the remaining balls. If the third ball is white, the game is over. If the third ball is red, the player wins another \$400.

(i) Draw a tree diagram showing the different outcomes of the game. [2]

(ii) Find the probability that the player wins nothing. Deduce the probability that the player draws at least one red ball. [3]

(iii) Find the probability that the player selects from Box B, given that the player wins \$300. [3]

9 It is known that the masses, in kilograms, of oranges and pears sold at a supermarket are normally distributed. The means and standard deviations of these distributions, and the selling prices, in \$ per kilogram, are shown in the table below.

	Mean (kg)	Standard deviation (kg)	Selling price (\$ per kg)
Oranges	0.27	c	\$3
Pears	0.39	0.05	\$6

(i) Given that 4% of the oranges has a mass less than 0.2 kg, show that c = 0.04, correct to 2 decimal places. [2]

(ii) Find the probability that a randomly chosen orange has mass greater than 340g. [1]

(iii) Find the probability that the mass of 3 randomly chosen oranges is within 0.01 kg of the mass of 2 randomly chosen pears, stating clearly the mean and variance of the distribution that you use. [3]

(iv) Find the probability that the cost of 3 randomly chosen oranges and 2 randomly chosen pears exceeds \$7, stating clearly the mean and variance of the distribution that you use. [3]

(v) State an assumption needed for the calculations in part **(iii)** and **(iv)** to be valid. [1]

10 In an egg farm, eggs are packed in cartons containing 30 eggs each. On average, 5% of the eggs are cracked during the transportation process from the egg farm to a market. At the market, every carton of 30 eggs is checked for cracked eggs. The number of cracked eggs in a randomly chosen carton is denoted by the random variable X.

(i) State, in the context of this question, two assumptions needed to model X using a binomial distribution. [2]

(ii) Explain why one of the assumptions stated in part **(i)** may not hold in this context. [1]

Assume now that these assumptions do in fact hold.

(iii) A carton is rejected if there is more than one cracked egg. Find the probability that a randomly chosen carton is rejected. [2]

(iv) 10 randomly chosen cartons of eggs are checked for cracked eggs. Find the probability that the last carton is the third rejected carton. [4]

(v) The eggs are also packed in trays of n eggs. Find the least value of n such that the probability of obtaining at most two cracked eggs in a randomly chosen tray is less than 0.99. [2]

11 An electric heater was switched on in a cold room and the temperature of the room was noted at five-minute intervals.

Time from switching on electric heater, x (min)	0	5	10	15	20	25	30	35	40
Temperature of room, y (°C)	0.4	1.5	3.4	5.5	7.7	9.7	11.7	13.5	15.4

(i) Draw a sketch of the scatter diagram for the data, as shown on your calculator. [2]

(ii) Find the product moment correlation coefficient and comment on its value in the context of the data. [2]

(iii) Find the equation of the regression line of y on x in the form $y = mx + c$, giving the values of m and c correct to 5 decimal places. Draw the line on the scatter diagram in part **(i)** and give an interpretation of m in the context of the question. [3]

(iv) Predict the temperature 2 hours from switching on the electric heater. Give a reason why should this prediction be treated with caution in the context of the question. [2]

(v) It was later found that the temperature was in fact k °C after the electric heater was switched on for 30 minutes, and the equation of the correct regression line of y on x should be $y = 0.4x$. Find the value of k. [2]

12 A parent claims that the average speed of vehicles along the road outside a particular school is greater than the speed limit of 40 km per hour. The Traffic Police recorded the speed, x km per hour, of 50 randomly selected vehicles along the road outside the school to obtain unbiased estimates of the population mean and variance of the speed. The data collected are summarised as follows.

$$\sum(x-40)=41,\ \sum(x-40)^2=5173.$$

(i) Suggest why, in this context, the data is summarised in terms of $(x-40)$ rather than x? [1]

(ii) Find unbiased estimates of the population mean and population variance. [3]

(iii) Test, at the 5% level of significance, whether there is sufficient evidence to support the parent's claim. [4]

(iv) State, with a reason, whether it is necessary to assume a normal distribution for the test in part **(iii)** to be valid. [1]

From past records, it is known that the speed along the road outside the school follows a normal distribution with standard deviation of 10 km per hour. To further investigate the parent's claim, the Traffic Police recorded the speed of another 20 randomly selected vehicles along the road outside the school and the mean speed for the second sample is c km per hour.

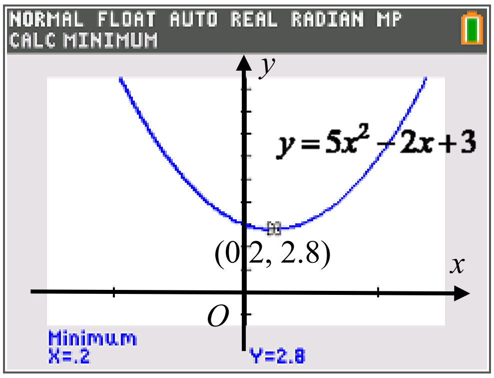
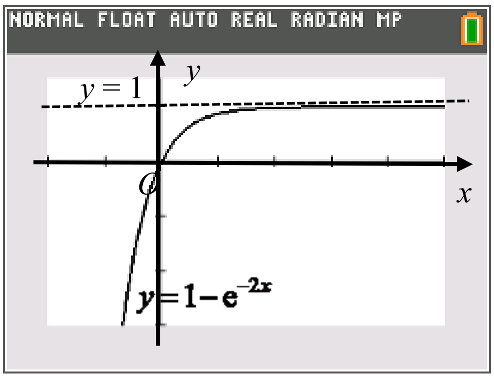
(v) Show that the unbiased estimate of the population mean speed based on the combined sample of 70 readings is given by $\dfrac{2041+20c}{70}$. [2]

(vi) Find the range of values of c such that there is sufficient evidence to support the parent's claim at the 5% level of significance, based on the combined sample. [3]

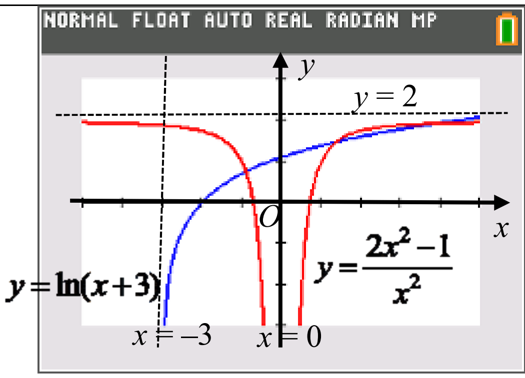
End of Paper

Millennia Institute
H1 Mathematics 2017 Prelim Exam Solution

1	Find the range of values of k for which the equation $x^2+2kx+3k+4=0$ has real roots. [3]
	Solution: $x^2+2kx+3k+4=0$ has real roots $\Rightarrow (2k)^2-4(1)(3k+4)\geq 0$ $\Rightarrow 4k^2-12k-16\geq 0$ $\Rightarrow k^2-3k-4\geq 0$ $\Rightarrow (k-4)(k+1)\geq 0$ $\Rightarrow k\leq -1$ or $k\geq 4$
2	Find $\frac{\mathrm{d}}{\mathrm{d}x}\left[\ln\left(\frac{\sqrt{2x^2+1}}{2-x}\right)\right]$. [3]
	Solution: $\frac{\mathrm{d}}{\mathrm{d}x}\left[\ln\left(\frac{\sqrt{2x^2+1}}{2-x}\right)\right]=\frac{\mathrm{d}}{\mathrm{d}x}\left[\frac{1}{2}\ln\left(2x^2+1\right)-\ln\left(2-x\right)\right]$ $=\frac{1}{2}\frac{1}{2x^2+1}\left(4x\right)-\frac{1}{2-x}(-1)$ $=\frac{2x}{2x^2+1}+\frac{1}{2-x}.$
3	The graph of $y=ax^2+bx+c$ passes through the points (1, 6), (7, 234) and (13, 822). By forming a system of linear equations, find the values of a, b and c. Hence find the range of values of x for which the graph is decreasing. [2]
	Solution: $y=ax^2+bx+c$ At (1, 6), $6=a+b+c$ … (1) At (7, 234), $234=49a+7b+c$ … (2) At (13, 822), $822=169a+13b+c$ … (3) From graphing calculator, a = 5, b = –2 and c = 3. **Method 1**: Differentiation $y=5x^2-2x+3\Rightarrow\frac{\mathrm{d}y}{\mathrm{d}x}=10x-2$ For y to be decreasing, $\frac{\mathrm{d}y}{\mathrm{d}x}<0\Rightarrow 10x-2<0\Rightarrow x<\frac{1}{5}.$

	Method 2: Complete the square $y=5\left(x^2-\frac{2}{5}x+\frac{3}{5}\right)=5\left[\left(x-\frac{1}{5}\right)^2-\left(-\frac{1}{5}\right)^2+\frac{3}{5}\right]$ $y=5\left[\left(x-\frac{1}{5}\right)^2+\frac{14}{25}\right]$ has a minimum point at $\left(\frac{1}{5},\frac{14}{25}\right)$ since a = 5 > 0. Hence, graph is decreasing when $x<\frac{1}{5}$. **Method 3**: Draw graph  Hence, graph is decreasing when $x<\frac{1}{5}$.
4	The curve C has equation $y=1-e^{-2x}$. **(i)** Sketch the graph of C, stating the equation(s) of any asymptote(s). **(ii)** Find the equation of the tangent to the curve C at $x=1$, giving your answer in the form $y=mx+c$, where m and c are exact constants to be found. **(iii)** Find the exact area bounded by the curve C, the line x = 1 and the x-axis. Deduce the exact area bounded by the curve C, the line y = $1-e^{-2}$ and the y-axis.
	Solution: (i)  (ii)

	$y=1-e^{-2x} \Rightarrow \dfrac{dy}{dx}=-e^{-2x}(-2)=2e^{-2x}$. When $x = 1$, $y=1-e^{-2}$ and $\dfrac{dy}{dx}=2e^{-2}$. **Method 1**: Use $y = mx + c$ $(1-e^{-2})=(2e^{-2})(1)+c \Rightarrow c=1-3e^{-2}$. Required equation is $y=2e^{-2}x+1-3e^{-2}$. **Method 2**: Use $y - b = m(x - a)$ Required equation is $y-\left(1-e^{-2}\right)=\left(2e^{-2}\right)\left(x-1\right)$ i.e. $y=2e^{-2}x-2e^{-2}+1-e^{-2}$ i.e. $y=2e^{-2}x+1-3e^{-2}$, i.e. $m=2e^{-2}$ and $c=1-3e^{-2}$. (iii) Required area = $\int_0^1\left(1-e^{-2x}\right)dx=\left[x-\dfrac{e^{-2x}}{-2}\right]_0^1$ $=\left(1+\dfrac{1}{2}e^{-2}\right)-\left(0+\dfrac{1}{2}\right)=\dfrac{1}{2}\left(1+e^{-2}\right)$. Required area = $\left(1-e^{-2}\right)(1)-\dfrac{1}{2}\left(1+e^{-2}\right)$ $=\dfrac{1}{2}-\dfrac{3}{2}e^{-2}=\dfrac{1}{2}\left(1-3e^{-2}\right)$. (deduced)
5	**(i)** Sketch, on the same diagram, the graphs of $y=\ln(x+3)$ and $y=\dfrac{2x^2-1}{x^2}$, labelling clearly the equations of any asymptotes. There is no need to find the coordinates of any points where the graphs cross the axes. **(ii)** Find the x-coordinate(s) of the point(s) of intersection of the graphs of $y=\ln(x+3)$ and $y=\dfrac{2x^2-1}{x^2}$. Hence solve the inequality $1+x^2\ln(x+3)<2x^2$. **(iii)** Find the area bounded by the two curves $y=\ln(x+3)$ and $y=\dfrac{2x^2-1}{x^2}$, giving your answer to 3 decimal places.
	Solution: (i)



(ii)
From graphing calculator,
Required coordinates are –0.893, 1.38, 3.92. (3 s.f.)

$1+x^2\ln(x+3)<2x^2$

$\Rightarrow \ln(x+3)<\dfrac{2x^2-1}{x^2}$ (since $x\neq 0$)

From graph, $-3<x<-0.893$ or $1.38<x<3.92$. (3 s.f.)

(iii)

Required area = $\displaystyle\int_{1.3841}^{3.9246}\left(\frac{2x^2-1}{x^2}-\ln(x+3)\right)\mathrm{d}x$ [from (ii)]

= 0.234 units^2. (3 d.p.) (from graphing calculator)

6 The managing director of a company tracked the rate of output, x units per month, of its product regularly over t months. His analyst believes that x and t can be modelled by the equation

$$x=a+30t^2-2t^3\text{, where } 0\le t\le 12 \text{ and } a \text{ is a positive constant.}$$

(i) Using differentiation, find the maximum rate of output in the year in terms of a, justifying that this is a maximum.

(ii) Sketch the graph of x against t for a = 25 and give an interpretation of the value of a. [3]

(iii) Find the exact area of the region bounded by the graph, the axes and the line t = 12 for a = 25. Give an interpretation of the value of this area.

The analyst also believes that the profit per month, \$$y$ million, can be modelled by the equation

$$y=\ln[(x-a)^2+(x-a)]\text{, where } x>a.$$

(iv) By expressing y in terms of t, find the rate at which the profit per month is increasing when t = 1.

Solution:

(i) $x=a+30t^2-2t^3$

At stationary point, $\dfrac{\mathrm{d}x}{\mathrm{d}t}=60t-6t^2=0$

$\Rightarrow 6t(10-t)=0 \Rightarrow t=0$ or $t=10$.

Method 1 Second derivative test

$\frac{d^2x}{dt^2} = 60 - 12t.$

When $t = 0$, $\frac{d^2x}{dt^2} = 60 > 0 \Rightarrow$ Rate is a minimum.

When $t = 10$, $\frac{d^2x}{dt^2} = -60 < 0 \Rightarrow$ Rate is a maximum.

Method 2 First derivative test

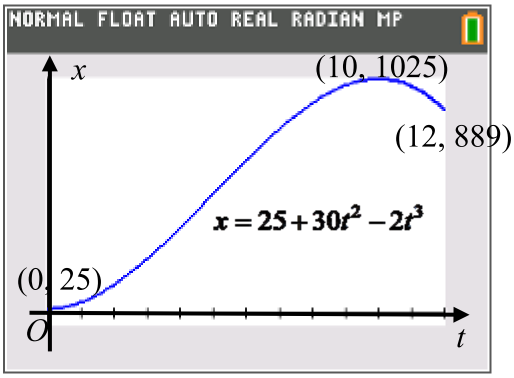
$\frac{dx}{dt} = 60t - 6t^2 = 6t(10 - t).$

t	0^-	0	0^+	t	10^-	10	10^+
$\frac{dx}{dt}$	< 0	0	> 0	$\frac{dx}{dt}$	> 0	0	< 0
Slope	\	—	/	Slope	/	—	\

Rate is a minimum at $t = 0$ and maximum at $t = 10$.

Required rate = $a + 30(10)^2 - 2(10)^3 = a + 1000.$

(ii)



The value of a represents the initial rate of output at the start of the year.

(iii) Required area

$= \int_0^{12} \left(25 + 30t^2 - 2t^3\right) dt = \left[25t + \frac{30t^3}{3} - \frac{2t^4}{4}\right]_0^{12}$

$= 25(12) + 10(12)^3 - \frac{1}{2}(12)^4 - 0 = 7212.$

It represents the yearly output, i.e. the company produces 7212 units in the year.

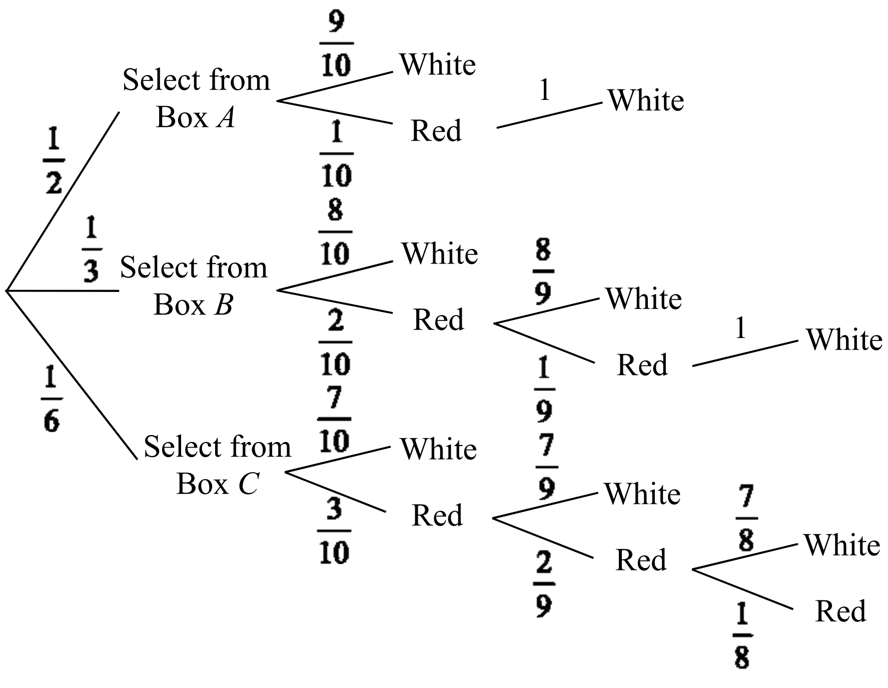
(iv)

$y = \ln[(x-a)^2 + (x-a)]$

$y = \ln[(30t^2 - 2t^3)^2 + (30t^2 - 2t^3)]$

From graphing calculator,

When $t = 1$, $\frac{dy}{dt} \approx 3.7906 = 3.79.$ (3 s.f.)

	Required rate = 3.79. (3 s.f.) .
7	Consider arranging all the letters of the word **FORMULAE**. **(i)** Find the number of different arrangements if there are no restrictions. **(ii)** Find the probability that the arrangement starts and ends with a consonant and the vowels are together. Codewords are formed by arranging 3 letters from the letters of the word **FORMULAE**. **(iii)** Find the number of different codewords that can be formed.
	(i) Number of ways = 8! = 40320. (ii) Required probability = $\frac{4\times 3\times 3!\times 4!}{40320}=\frac{1728}{40320}=\frac{3}{70}$. (iii) Number of codewords = ${}^{8}C_{3}\times 3! = 336$.
8	In a game, there are three boxes *A*, *B* and *C*. Box *A* contains 1 red and 9 white balls. Box *B* contains 2 red and 8 white balls. Box *C* contains 3 red and 7 white balls. All the red and white balls in the three boxes are indistinguishable other than the colours. The player selects one of the three boxes by tossing a fair die. The player selects a ball fr If the player selects a white ball, the game is over. If the player selects a red ball, the pers the same box containing the remaining balls. If the third ball is white, the game is over. If the third ball is red, the player wins another $400. **(i)** Draw a tree diagram showing the different outcomes of the game. **(ii)** Find the probability that the player wins nothing. Deduce the probability that the player draws at least one red ball. **(iii)** Find the probability that the player selects from Box *B*, given that the player wins $300. [3]
	(i)  (ii)

	P(wins nothing) = $\frac{1}{2}\times\frac{9}{10}+\frac{1}{3}\times\frac{8}{10}+\frac{1}{6}\times\frac{7}{10}=\frac{50}{60}=\frac{5}{6}$. P(draws ≥ 1 red ball) = P(wins something) $=1-\text{P(wins nothing)} = 1-\frac{5}{6}=\frac{1}{6}$. (iii) P(selects from Box B | wins \$300) $=\dfrac{\text{P(selects from Box } B \text{ and wins \$300)}}{\text{P(wins \$300)}}$ $=\dfrac{\frac{1}{3}\times\frac{2}{10}\times\frac{1}{9}}{\frac{1}{3}\times\frac{2}{10}\times\frac{1}{9}+\frac{1}{6}\times\frac{3}{10}\times\frac{2}{9}\times\frac{7}{8}}=\dfrac{\frac{2}{3}}{\frac{2}{3}+\frac{7}{8}}=\frac{16}{37}$.
9	It is known that the masses, in kilograms, of oranges and pears sold at a supermarket are normally distributed. The means and standard deviations of these distributions, and the selling prices, in \$ per kilogram, are shown in the table below. | | Mean (kg) | Standard deviation (kg) | Selling price (\$ per kg) | | Oranges | 0.27 | c | \$3 | | Pears | 0.39 | 0.05 | \$6 | **(i)** Given that 4% of the oranges has a mass less than 0.2 kg, show that c = 0.04, correct to decimal places. [2] **(ii)** Find the probability that a randomly chosen orange has mass greater than 340g.[1] **(iii)** Find the probability that the mass of 3 randomly chosen oranges is within 0.01 kg of the mass of 2 randomly chosen pears, stating clearly the mean and variance of the distribution that you use. [3] **(iv)** Find the probability that the cost of 3 randomly chosen oranges and 2 randomly chosen pears exceeds \$7, stating clearly the mean and variance of the distribution that you use. [3] **(v)** State an assumption needed for the calculations in part **(iii)** and **(iv)** to be valid. [1]
	Solution: (i) Let X be the mass of a randomly chosen orange. Then $X \sim \text{N}(0.27, c^2)$. P($X$ < 0.2) = 0.04. $\Rightarrow \text{P}\left(Z<\frac{0.2-0.27}{c}\right)=0.04$, where $Z \sim \text{N}(0, 1)$ From graphing calculator, $\frac{-0.07}{c}\approx -1.7507$ $\Rightarrow c\approx 0.03998 = 0.04$. (2 d.p.) (shown) (ii) From graphing calculator, P(X > 0.34) ≈ 0.040059 = 0.0401. (3 s.f.) OR Required probability = P(X > 0.34) = P(X < 0.2) = 0.04. (by symmetry)

	(iii) Let Y and W be the mass, in kg, of 3 randomly chosen oranges and 2 randomly chosen pears respectively. Then $Y \sim N(3\times0.27,\ 3\times0.04^2)$, i.e. N(0.81, 0.0048) and $W \sim N(2\times0.39,\ 2\times0.05^2)$, i.e. N(0.78, 0.005). $Y - W \sim N(0.81 - 0.78,\ 0.0048 + 0.005)$, i.e. N(0.03, 0.0098). Required probability $= P(W - 0.01 < Y < W + 0.01)$ $= P(-0.01 < Y - W < 0.01)$ $\approx 0.076862 = 0.0769$. (3 s.f.) (iv) Let S and T be the cost, in \$, of 3 randomly chosen oranges and 2 randomly chosen pears respectively. Then $S = 3Y \sim N(3\times0.81,\ 3^2\times0.0048)$, i.e. N(2.43, 0.0432) and $T = 6W \sim N(6\times0.78,\ 6^2\times0.005)$, i.e. N(4.68, 0.18). $S + T \sim N(2.43 + 4.68,\ 0.0432 + 0.18)$, i.e. N(7.11, 0.2232). Required probability $= P(S + T > 7)$ $\approx 0.59205 = 0.592$. (3 s.f.) (v) We need to assume that the masses of oranges and pears are independent.
10	In an egg farm, eggs are packed in cartons containing 30 eggs each. On average, 5% of the eggs are cracked during the transportation process from the egg farm to a market. At the market, every carton of 30 eggs is checked for cracked eggs. The number of cracked eggs in a randomly chosen carton is denoted by the random variable X. **(i)** State, in the context of this question, two assumptions needed to model X using a binomial distribution. **(ii)** Explain why one of the assumptions stated in part **(i)** may not hold in this context. [1] Assume now that these assumptions do in fact hold. **(iii)** A carton is rejected if there is more than one cracked egg. Find the probability that a randomly chosen carton is rejected. **(iv)** 10 randomly chosen cartons of eggs are checked for cracked eggs. Find the probability that the last carton is the third rejected carton. **(v)** The eggs are also packed in trays of n eggs. Find the least value of n such that the probability of obtaining at most two cracked eggs in a randomly chosen tray is less than 0.99.
	Solution: (i) We need to assume that: 1. the event that an egg is cracked is independent of that of other eggs. 2. the probability that an egg is cracked is a constant. (ii)

In this context, the event that an egg is cracked (due to transportation) may affect neighbouring eggs in the same carton to crack and thus the events may not be independent.

(iii)

P(reject carton) = $P(X > 1)$ where $X \sim B(30, 0.05)$

$= 1 - P(X \leq 1)$

$\approx 0.44646 = 0.446$. (3 s.f.)

(iv)

Let Y be the number of rejected cartons, out of 9.

Then $Y \sim B(9, 0.44646)$.

Required probability = $P(Y = 2) \times 0.44646$

$\approx 0.051015 = 0.0510$. (3 s.f.)

(v)

Let W be the number of cracked eggs in a tray, out of n.

Then $W \sim B(n, 0.05)$.

We want to find least n such that $P(W \leq 2) < 0.99$.

n	$P(W \leq 2)$
9	$0.99164 > 0.99$
10	$0.9885 < 0.99$
11	$0.98476 < 0.99$

From graphing calculator, least value of n is 10.

11 An electric heater was switched on in a cold room and the temperature of the room was noted at five-minute intervals.

Time from switching on electric heater, x (min)	0	5	10	15	20	25	30	35	40
Temperature of room, y (°C)	0.4	1.5	3.4	5.5	7.7	9.7	11.7	13.5	15.4

(i) Draw a sketch of the scatter diagram for the data, as shown on your calculator.

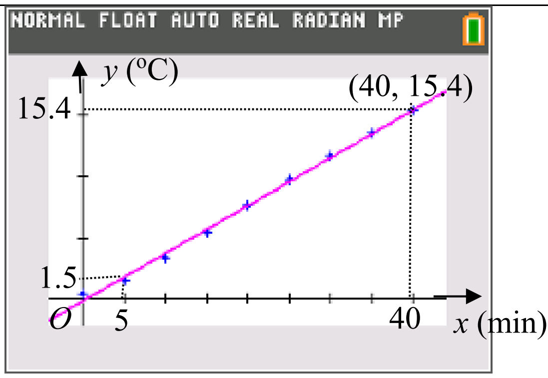
(ii) Find the product moment correlation coefficient and comment on its value in the context of the data.

(iii) Find the equation of the regression line of y on x in the form $y = mx + c$, giving the values of m and c correct to 5 decimal places. Draw the line on the scatter diagram in part **(i)** and give an interpretation of m in the context of the question.

(iv) Predict the temperature 2 hours from switching on the electric heater. Give a reason why should this prediction be treated with caution in the context of the question.

(v) It was later found that the temperature was in fact k °C after the electric heater was switched on for 30 minutes, and the correct regression line should be $y = 0.4x$. Find the value of k.

(i)

	 (ii) From graphing calculator, required coefficient, $r \approx 0.99870 = 0.999$. (3 s.f) Since $r > 0$ and $\lvert r\rvert = 0.999$ is very close to 1, the value of r suggests that there is a strong positive linear correlation between the temperature of the room and the time from switching on the electric heater. (iii) From graphing calculator, required equation is $y = 0.38933x - 0.14222$. (5 d.p) The temperature of the room is expected to increase by 0.3893^{O}C for every one minute increase in time from switching on the electric heater. (iv) When $x = 120$, $y = 0.38933(120) - 0.14222 \approx 46.577 = 46.6$. (3 s.f.) Required temperature is 46.6 °C. Possible reasons why prediction should be treated with caution: 1. Extrapolation This prediction of 46.6^{O}C is rather high and should be treated with caution since $x = 60$ is far outside the range of values of x in the data, i.e. $y = 46.6$ is an extrapolation. 2. Relationship not linear outside data range. The relationship between the temperature of the room and the time from switching on the electric heater may not be linear any more beyond 40 minutes. 3. Actual temperature too high and can cause a fire. The actual temperature 120 minutes after switching on the electric heater may be high enough to cause a fire which can be a disaster. (v) For the new data, $\sum x = 180$, $\sum y = 57.1 + k$. $y = 0.4x \Rightarrow \bar{y} = 0.4\bar{x} \Rightarrow \sum y = 0.4\sum x$ $\Rightarrow 57.1 + k = 0.4(180) \Rightarrow k = 14.9$.
12	A parent claims that the average speed of vehicles along the road outside a particular school is greater than the speed limit of 40 km per hour. The Traffic Police recorded the speed, x km per hour, of 50 randomly selected vehicles along the road outside the school to obtain unbiased estimates of the population mean and variance of the speed. The data collected are summarised as follows. $$\sum(x-40) = 41, \ \sum(x-40)^2 = 5173.$$

(i) Suggest why, in this context, the data is summarised in terms of $(x-40)$ rather than x?

(ii) Find unbiased estimates of the population mean and population va

(iii) Test, at the 5% level of significance, whether there is sufficient evidence to support t

(iv) State, with a reason, whether it is necessary to assume a normal distribution for the t

From past records, it is known that the speed along the road outside the school follows a normal distribution with standard deviation of 10 km per hour. To further investigate the parent's claim, the Traffic Police recorded the speed of another 20 randomly selected vehicles along the road outside the school and the mean speed for the second sample is c km per hour.

(v) Show that the unbiased estimate of the population mean speed based on the combined sample of 70 readings is given by $\dfrac{2041+20c}{70}$.

(vi) Find the range of values of c such that there is sufficient evidence to support the parent's claim at the 5% level of significance, based on the combined sample.

Solution:

(i) This is to keep the recorded speed values small or to give an indication of the variations around the hypothesised mean speed of 40km/h.

(ii)

Let $y = x - 40$. Then $\sum y = 41$, $\sum y^2 = 5173$.

Unbiased estimate of population mean,

$$\overline{x} = \overline{y} + 40 = \frac{\sum y}{50} + 40 = \frac{41}{50} + 40 = 40.82.$$

Unbiased estimate of population variance,

$$s_x^2 = s_y^2 = \frac{1}{50-1}\left(\sum y^2 - \frac{1}{50}\left(\sum y\right)^2\right)$$

$$= \frac{1}{49}\left(5173 - \frac{41^2}{50}\right) = \frac{5139.38}{49} \approx 104.89 = 105. \text{ (3 s.f.)}$$

(iii)

Let X be the speed of a randomly chosen vehicle along the road outside the particular school.

Test $H_0 : \mu = 40$ against $H_1 : \mu > 40$ (claim)

Under H_0, since $n = 50$ is large, by Central Limit Theorem, $\overline{X} \sim N(40, \frac{104.89}{50})$ approximately.

Using a one-tail z-test, p-value ≈ 0.28564.

Since p-value $\approx 0.28564 > 0.05$, we do not reject H_0 at the 5% level of significance and conclude that there is not enough evidence to support the parent's claim at the 5% level of significance.

(iv)

It is not necessary to assume a normal distribution for the test in part (iii) to be valid

	since $n = 50$ is large, by Central Limit Theorem, $\overline{X}$ is normally distributed approximately. (v) Required estimate = $\dfrac{\sum x + 20c}{50+20} = \dfrac{50\overline{x}+20c}{70} = \dfrac{2041+20c}{70}$. (shown) (vi) Test $H_0 : \mu = 40$ against $H_1 : \mu > 40$ (claim) Under H_0, $\overline{X} \sim N\left(40, \dfrac{10^2}{70}\right)$, i.e. $N\left(40, \dfrac{10}{7}\right)$. Using a one-tail z-test at $\alpha = 0.05$, critical value = 41.966. (from graphing calculator) To have sufficient evidence to support the claim at 5% level of significance, we reject H_0 at $\alpha = 0.05$. $\dfrac{2041+20c}{70} \geq 41.966$ i.e. $c \geq 44.8$. (3 s.f.)