CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

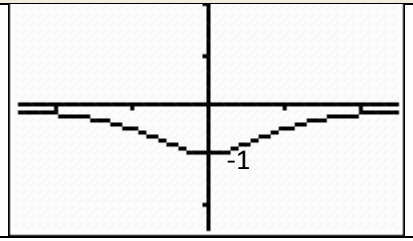
1	**SLE**		
Part	**Assessment Objectives**	**Mark Scheme**	
	Able to form system of linear equations Able to solve a system of linear equations using GC	Let $f(x) = ax^3 + bx^2 + cx + d$. Also, $f'(x) = 3ax^2 + 2bx + c$ $f(2) = 8a + 4b + 2c + d = 6$ $f(-1) = -a + b - c + d = 15$ $f'(1) = 3a + 2b + c = 0$ $f'(-2) = 12a - 4b + c = 0$ Solving, $a = 2, b = 3, c = -12, d = 2 \Rightarrow f(x) = 2x^3 + 3x^2 - 12x + 2$	

2	**Vectors**		
Part	**Assessment Objectives**	**Mark Scheme**	
(i)	Find intersection between line and plane	At pt of intersection, $\begin{pmatrix} 1+3\lambda \\ 1-\lambda \\ -2+2\lambda \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} = 4$ $\Rightarrow \lambda = 9$ So intersection pt is (28, -8, 16).	
(ii)	Form equation of plane given a line and a point	A direction vector on p_2 is $\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} - \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$	

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

2	**Vectors**		
Part	**Assessment Objectives**	**Mark Scheme**	
		A vector normal to $p2$ is $\begin{pmatrix}1\\2\\0\end{pmatrix}\times\begin{pmatrix}3\\-1\\2\end{pmatrix}=\begin{pmatrix}4\\-2\\-7\end{pmatrix}$ So an equation of $p2$ is $\mathbf{r}\cdot\begin{pmatrix}4\\-2\\-7\end{pmatrix}=\begin{pmatrix}0\\-1\\-2\end{pmatrix}\cdot\begin{pmatrix}4\\-2\\-7\end{pmatrix}=16$ i.e. $\mathbf{r}\cdot\begin{pmatrix}4\\-2\\-7\end{pmatrix}=16$	
(iii)	Find intersection of 2 planes using GC	$-x+2z=4$ and $4x-2y-7z=16$. Solving, an equation of line is $\mathbf{r}=\begin{pmatrix}-4\\-8\\0\end{pmatrix}+\mu\begin{pmatrix}4\\1\\2\end{pmatrix}$, $\mu\in\Re$	
(iv)	Understand relationship between normal of a plane and a line // to it	For p_3 // l_2, $\begin{pmatrix}1\\a\\b\end{pmatrix}\cdot\begin{pmatrix}4\\1\\2\end{pmatrix}=0$ $\Rightarrow a+2b+4=0$	

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

3	**FUNCTIONS**		
Part	**Assessment Objectives**	**Mark Scheme**	
(i)	Draw graph of a function		-
(ii)	Understand condition for inverse to exist	**For inverse to exist, function f must be one-one** and thus, **largest k is 0.**	
(iii)	• Know how to find inverse function and select the appropriate function depending on domain of f • Understanding relationship between f and its inverse HOT	$y = -\frac{1}{x^2+1}$ $x^2 y + y = -1$ $x^2 = \frac{-1-y}{y}$ $x = \pm\sqrt{\frac{-1-y}{y}}$ $f^{-1}(x) = -\sqrt{\frac{-1-x}{x}}$ Solving $f^{-1}(x) = f(x)$ is equivalent to solving $f(x) = x$: $-\frac{1}{x^2+1} = x$ $x^3 + x + 1 = 0$ Solving and taking the real root: $a = -0.682$ (3sf)	
(iv)	• Understand the condition for composite function to exist • Know how to restrict domain of a function to	For composite function to exist, $R_g \subseteq D_{f^{-1}} = [-1,0)$ Restricted domain of g $= [-4, -3)$	

3	**FUNCTIONS**		
Part	**Assessment Objectives**	**Mark Scheme**	
	achieve required range HOT		

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

4	**Definite Integral**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i)	Know how and when to use double angle formula Know how to apply integration by parts	$\int x\sin^2 x\,\mathrm{d}x$ $= \int x\left(\frac{1-\cos 2x}{2}\right)\mathrm{d}x$ $= \frac{1}{2}\int x\,\mathrm{d}x - \frac{1}{2}\int x(\cos 2x)\,\mathrm{d}x$ $= \frac{1}{2}\left[\frac{x^2}{2}\right] - \frac{1}{2}\int x(\cos 2x)\,\mathrm{d}x$ $u = x \quad \frac{\mathrm{d}v}{\mathrm{d}x} = \cos 2x$ $\frac{\mathrm{d}u}{\mathrm{d}x} = 1 \quad v = \frac{\sin 2x}{2}$ $\int x^2\sin^2 x\,\mathrm{d}x = \frac{1}{2}\left[\frac{x^2}{2}\right] - \frac{1}{2}\left[x\frac{\sin 2x}{2} - \frac{1}{2}\int \sin 2x\,\mathrm{d}x\right]$ $= \frac{1}{4}\left[x^2 - x\sin 2x - \frac{1}{2}\cos 2x\right] + \mathrm{C}$	Less than half the student cohort correctly used the double angle as part of the problem solving procedure. Only a handful of students used the alternative methods and lead to a longer solution. They were conceptually correct and many did well.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

4	**Definite Integral**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(ii)	Know how to use definite integral to find volume of solid of revolution Know how to use previous answer in evaluation of definite integral	$V = \pi\int_0^{\pi} x\sin^2 x\,\mathrm{d}x$ $V = \frac{\pi}{4}\left[x^2 - x\sin 2x - \frac{1}{2}\cos 2x\right]_0^{\pi}$ $V = \frac{\pi}{4}\left[\pi^2 - \pi\sin 2\pi - \frac{1}{2}\cos 2\pi + \frac{1}{2}\right] = 7.75$	This part is generally well done.
(iii)	Know how to find required volume by using volume of previous solid found	$\text{Volume} = \pi\left(\sqrt{\pi}\right)^2(\pi) - 7.7516$ $= 23.3$	Most students could not even state the volume of cylinder correctly. Some do not even know how to subtract the volume from one another. This part is badly done. I suspect the students don't have a visual approach to solving this part. They need to use a diagram to aid their problem solving.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

5	DIFFERENTIAL EQUATIONS AND APPLICATIONS		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
	• Able to solve DE using substitution	$xy\dfrac{\mathrm{d}y}{\mathrm{d}x} = x^2 + y^2$ $y = ux$ $\dfrac{\mathrm{d}y}{\mathrm{d}x} = u + x\dfrac{\mathrm{d}u}{\mathrm{d}x}$ $x(ux)\left(u + x\dfrac{\mathrm{d}u}{\mathrm{d}x}\right) = x^2 + (ux)^2$ $u^2x^2 + ux^3\dfrac{\mathrm{d}u}{\mathrm{d}x} = x^2 + (ux)^2$ $\dfrac{\mathrm{d}u}{\mathrm{d}x} = \dfrac{1}{ux}$ $u\dfrac{\mathrm{d}u}{\mathrm{d}x} = \dfrac{1}{x}$ $\int u\,\mathrm{d}u = \int \dfrac{1}{x}\,\mathrm{d}x$ $\dfrac{u^2}{2} = \ln\lvert x\rvert + c$ $u^2 = 2\ln\lvert x\rvert + d$ $y^2 = 2x^2\ln\lvert x\rvert + dx^2$	This part is generally well done. A handful of students forgot to include the modulus sign.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

5	DIFFERENTIAL EQUATIONS AND APPLICATIONS		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
	HOT • Able to find particular solution, given conditions •	For the curve that passes (1, 2): $2^2 = 2(1)^2 \ln\|1\| + d(1)^2$ $d = 4$ $y^2 = 2x^2 \ln\|x\| + 4x^2$ When y = 3 (from GC): x = 1.39, – 1.39	Quite a number of students stated 1.39 as the only value for x.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

6	PERMUTATION AND COMBINATION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i) HOT	Solve counting problems by applying the idea of combinations.	Total possible groupings $= {}^{9}C_{2} \times {}^{29}C_{2} \times {}^{27}C_{2}$ $= 5130216$	The question was very badly done with many kinds of mistakes including using permutation instead, adding instead of multiplying and also many who simply gave ${}^{9}C_{2} \times {}^{29}C_{4}$, not realising that the 4 not involved in DI must still be split into pairs.
(ii) HOT		Total possible groupings $= \left({}^{8}C_{1} \times {}^{29}C_{2} \times {}^{27}C_{2}\right) + \left({}^{8}C_{2} \times {}^{28}C_{1} \times {}^{27}C_{2}\right) \times 2$ $= 1690416$	Out of those who completed (i) correctly, only a few went on to get this part correct. The best solutions used the idea of complement instead. Many made the error of using $\left({}^{9}C_{2} \times {}^{28}C_{1} \times {}^{27}C_{2}\right)$ instead, not realising that Jill cannot be among the ones selected for DI component.
(iii) HOT		Total possible groupings $= \left({}^{9}C_{2} \times {}^{28}C_{1} \times {}^{27}C_{2}\right) \times 2$ $= 707616$	Out of those who completed (i) correctly, only a few went on to get this part correct. The best solutions used the idea of complement instead. Many made the error of using $\left({}^{9}C_{2} \times {}^{29}C_{1} \times {}^{27}C_{2}\right)$ instead, not realising that Jack cannot be among the ones selected for the component which they had assumed that Jack is already in.
(iv) HOT	Calculating probabilities using P&C. Relating probabilities to the idea of fairness.	From the previous parts, if the arrangement is at random, the probability of a CI student on duty $= \frac{1690416}{5130216} = 0.330$ the probability of a SS student on duty $= \frac{707616}{5130216} = 0.138$ Hence, it is unfair as the probability that a CI student is put on	This part was also badly attempted as most could not get the previous parts correct. Majority of the students realised that fairness has to do with equal chance and probability, but did not fulfil the 'hence' requirement, which is to use the arrangements to calculate probability.

6	PERMUTATION AND COMBINATION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
		duty is much higher than a SS student.	Arguments made on the number of students in each wing and/or the special requirement of the DI component gained no credit. Given that most students were unable to complete the first few parts, those who got values of (ii) greater than (iii) and goes on to compare the number of arrangements and hence the probability were given credit for the analysis.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

7	PROBABILITY		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i)	Setting up equations and solving for unknowns involving probabilities	$0.75\times[(0.20\times0.91)+(0.63\times0.72)+(0.17\times0.54)]+(0.25\times a\times0.36)=0.60$ $0.09a=0.05445 \quad\Rightarrow\quad a=0.605$	This part was well-attempted with students losing marks only due to lack of accuracy in calculations, especially 0.9 instead of 0.09. Some students used the longer way of considering P(Not Officer) which wasted a considerable amount of time.
(ii)	Calculating conditional probability	Prob required $=\dfrac{\text{Prob(Fail NPFA and Officer)}}{\text{Prob(Officer)}}$ $=\dfrac{0.25\times0.605\times0.36}{0.60}$ $=0.0908$ (3 s.f.)	Most students were able to identify the need for a conditional probability, but made errors with the numerator, such as omitting the 0.25.
(iii)	Using the idea of complement for conditional probabilities	Prob required $=1-0.0908=0.909$ (3 s.f.)	Although most students figured that the answer is related to part (ii), many used 0.6 instead of 1, failing to realise that the sum of complementing conditional probabilities should still add up to 1. Some students wasted time re-calculating the conditional probability from scratch. Those who realised that 0.05445 from part (i) is the numerator should be commended.
(iv)	Identifying the different phrases to indicate conditional probabilities HOT	Prob required $=\dfrac{\text{Prob(Fail NPFA and Non-officer)}}{\text{Prob(Non-officer)}}$ $=\dfrac{0.25\times0.395+0.25\times0.605\times0.64}{1-0.60}$ $=0.489$ (3 s.f.)	Many failed to understand this phrasing of a conditional probability and gave only the numerator instead. Others who identified the conditional probability made errors with the numerator, often omitting 0.25 x 0.395.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

8	BINOMIAL DISTRIBUTION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i)	Identifying the conditions for a binomial model to be applied	The number of SMSes that John receives in each 30 minute period is independent of the number of SMSes John received in other 30 minute periods. (i.e. independent trials) The probability that John receives at least one SMS in a 30 minute period remains constant for all such periods. (i.e. constant probability of success)	This part of the question proved to be difficult for many students. While many mentioned to some effect, 14 independent trials and constant probability of success, no mark was awarded as there was no context to their answers. Also, there were many inaccurate phrasing, such as the probability was independent, the period was independent. Some students also thought that the rv X denotes the no of periods that John receive a SMS, when clearly, the question states at least a SMS.
(ii)	Making sense of the real life context of mathematical models HOT	The people who contact John will do so at their convenient time, which should not be constant throughout the day.	There were a varied answers for this part of the question, however, full credit was only awared if the student's answer provided some evidence that the probability was not a constant across periods. Also, there was a need for some context understanding, and it is not enough to merely write, there was a differing number of SMS received, an explanation for why so is needed.
(iii)	Calculating probabilities for binomial distributions	$X \sim \text{B}(14, 0.95)$ $P(X \leq 10) = 0.00417$	This part was very well-done.
(iv)	Using Poisson to approximate a binomial distribution and find the mode of a distribution HOT	Let Y be the random variable number of periods where John receives no SMS out of 70. $Y \sim \text{B}(70, 0.05)$ Since $n = 70 > 50$ and $np = 70 \times 0.05 = 3.5 < 5$ $Y \sim \text{Po}(3.5)$ approx	Some students did not remember the conditions for which a normal approximation can be done, and so, erroneously, approximate this given situation to that of a normal distribution. Also, a significant number of students also

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

8	BINOMIAL DISTRIBUTION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
		From GC, $Y = 3$	did not seem to understand that the concept of mode, with many finding the value y for which $P(Y= y) = 0.05$. Some also assume since $E(Y) = 3.5$, the mode = 4 (rounding up of 3.5)

9	POISSON DISTRIBUTION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i)	Calculating probabilities for Poisson distribution	$X \sim \text{Po}(15)$ $\quad P(X \geq 10) = 0.930$	This part was well-done.
(ii)		$Y \sim \text{Po}(19)$ $\quad P(Y \geq 10) = 0.991$	This part was well-done.
(iii) HOT	Using Normal approximation for Poisson.	$A \sim \text{Po}(30)$ Since $\lambda = 30 > 10$ $A \sim \text{N}(30, 30)$ approx. $B \sim \text{Po}(38)$ Since $\lambda = 38 > 10$ $B \sim \text{N}(38, 38)$ approx. $C = 20 + A - B$ $\quad C \sim \text{N}(12, 68)$ approx. $P(C = 5) \xrightarrow{c.c.} P(4.5 < C < 5.5)$ $= 0.0377$	Although quite a significant many of students scored full credit for this part, a majority showed misunderstanding, such as • $A \sim \text{Po}(30)$, $B \sim \text{Po}(38)$, implies $B - A \sim \text{Po}(38-30)$. • Additionally, some students even went on to approximate $B - A \sim \text{N}(8,8)$ when clearly $\lambda < 10$.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

10	**SAMPLING METHODS**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i)	Know when a particular sampling method can be used	Quota Sampling. No sampling frame available.	
(ii)	Know how a method can be applied	Susan can choose 25 males and 25 females from the food centre to interview, out of which 10 people are aged between 13 to 20, 20 aged between 21 to 30 and 20 aged between 31 to 50.	

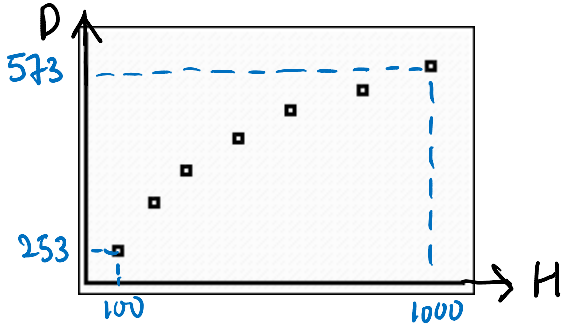
CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

11	NORMAL DISTRIBUTION + SAMPLING DISTRIBUTION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
	Understand condition for CLT and how to apply Apply concept of sampling distribution Know how to apply standardisation to solve problem	Let X be the random variable. Assuming that $n > 30$, by CLT, $\overline{X} \sim N(\mu, \frac{10}{n})$ approx $P(\mu - 0.5 < \overline{X} < \mu + 0.5) > 0.89$ $P\left(\frac{\mu - 0.5 - \mu}{\sqrt{\frac{10}{n}}} < Z < \frac{\mu + 0.5 - \mu}{\sqrt{\frac{10}{n}}}\right) > 0.89$ $P\left(-0.5\sqrt{\frac{n}{10}} < Z < 0.5\sqrt{\frac{n}{10}}\right) > 0.89$ $P\left(Z < -0.5\sqrt{\frac{n}{10}}\right) < 0.055$ $-0.5\sqrt{\frac{n}{10}} < -1.59819$ $\sqrt{\frac{n}{10}} > 3.19638$ $\frac{n}{10} > 10.21685$ $n > 102.1685$ Least $n = 103$	Alternative assumption is X follows normal distribution.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012

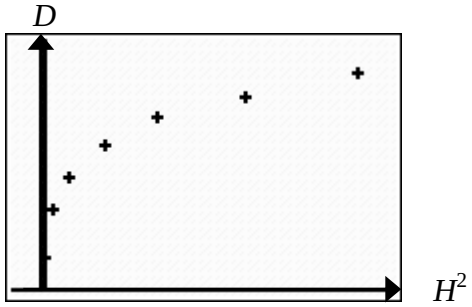
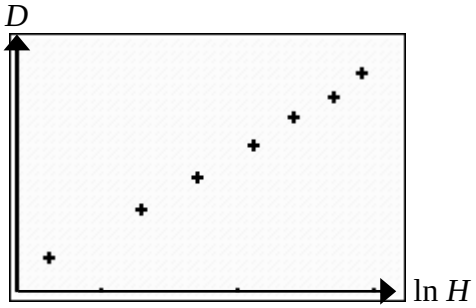
12	**HYPOTHESIS TESTING**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i)	- able to calculate the unbiased estimates of population mean and variance	$\bar{x} = \frac{\sum(x-500)}{10} + 500 = 580$ $s^2 = \frac{1}{9}\left[65764 - \frac{(800)^2}{10}\right] = 196$	The question was generally well done. Many students were able to calculate the unbiased estimates of the population mean and variance. A handful of students made arithmetic errors in the calculation of s^2.
(ii)	-able to formulate and carry out a 2-tail hypothesis test -use a t-test for a population mean using a small sample drawn from a normal population of unknown variance -understand that the p value is the smallest level of significance at which the null hypothesis can be rejected	Let X be the random variable denoting the mass of a cake. Assuming X is normally distributed, $\overline{X} \sim N\left(600, \frac{196}{10}\right)$ $H_0 : \mu = 600$ $H_1 : \mu \neq 600$ Level of significance = α Reject H_0 implies p value < α Test statistic, $t = \frac{580-600}{\sqrt{\frac{196}{10}}} = -4.52$ Degree of freedom = 9 p value = 0.00145 $\alpha > 0.00145$	This part was generally well done. Some students were penalised for the presentation in the writing of the hypotheses in the manner: $H_0 : \mu_0 = 600$ $\quad H_0 : \mu_0 = 600$ $H_1 : \mu_0 \neq 600$ or $H_1 : \mu_1 \neq 600$ There were a number of students who were confused about the difference in the use of μ_0 and $\bar{x}$. There were also students who made mistakes in the inequality sign, forgetting that the rejection of H_0, implies p value < α.
(iii)	- able to carry out a hypothesis test for a large sample drawn from any population, using the Central Limit Theorem and an unbiased estimate of the population variance - interpret the results of a hypothesis test in the context	$s^2 = \frac{100}{99}(33^2) = 1100$ Level of significance = 0.05 Since the bakery did not overstate the mean mass of the cakes, we do not reject H_0 p value > 0.05	The majority of the students were not aware that there is a difference between sample variance and unbiased estimate of population variance. Most of the students were not able to obtain s^2 accurately. A handful of students made errors in the formulae for the test statistic. Students who obtained least m = 595 will not be awarded the answer mark if he/she

12	HYPOTHESIS TESTING		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
	of the problem HOT	Test statistic, $z = \dfrac{m-600}{\sqrt{\frac{1100}{100}}}$ $\dfrac{m-600}{\sqrt{\frac{1100}{100}}} > -1.6449$ $m - 600 > -5.4554$ $m > 594.54$ Least $m = 595$ It is not necessary to assume that the mass of the cakes are normal distributed because CLT can be applied since sample size is large.	did not use the correct value for s^2. Students need to be clear in their answer on whether it is necessary or not necessary for the assumption that the mass of the cakes are normally distributed. The use of CLT is determined by sample size, not whether population variance is known/unknown.

13	CORRELATION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(i)	- To sketch scatter diagram	Plot of (horizontal dist. travelled / mm), D, against (height of release above the table / mm), H.  	This part of Q13 is generally well-done, with an occasional occurrence of mistakes, among which the more notable ones include : • an omission of axis-labels ; • a generic labelling of the horizontal axis as 'x' and the vertical axis as 'y', instead of specifically using 'H' and 'D' respectively for the horizontal and vertical axes ; • inaccurate positioning of data points.

13	**CORRELATION**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(ii)	- Understand the implication of r value - Interpreting relationship using scatter diagram	Using the example in this problem, even though the value of $r = 0.962\ldots$ may be such that $\lvert r\rvert$ is close to +1, which indicates that the set of data points on the scatter plot lie close to some straight line, the data points could be a lot more closely-fitted to a curved trend (along which variable D increases at a rate that decreases with respect to H as variable H increases) – – which is non-linear. As such, solely relying on the value of r to determine whether the relationship between two variables is linear can be insufficient in doing so conclusively.	This part of Q13 is generally not very well nor precisely explained. Of the set of answer-responses received for this part of Q13, quite a significant number did not adequately explain *why* the value of r alone is insufficient in indicating a linear relationship between two variables, simply pointing out that the points on the scatter diagram follows a curve without contrasting this observation with an interpretation of the sample correlation coefficient $r = 0.962\ldots$, which on the contrary is indicative of a strong linear relationship between the two relevant variables. Note: It would be helpful to recognise that giving an example in which "*a* happens to be true, but not *b*" is a demonstration of "*a* not being a sufficient condition for *b*". Within the set of responses that provided a contrast between the what the *r*-value indicates and what the scatter plot exhibits, some responses happened to be a little too general in simply stating that the points do not follow a straight line / there is no linear relationship as observed from the scatter plot (and this is a little too

13	CORRELATION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
			general because there are many instances of scatter plots with an *r*-value close to 1 that would also befit this general description, such one in which two clusters of points are present but positioned far-apart from each other.) Mentioning that there's a more closely-fitted curve / non-linear trend to the plotted data points on the scatter diagram as compared to any straight line, however, would be a much more precise way of description. A few of the responses attempted to identify the nature of the non-linear trend or relationship, sometimes erroneously describing the exhibited relationship between *D* and *H* as "exponential" or "quadratic", when a description of the exhibited trend as "curved" or "non-linear" would be adequate in providing the required contrast to explain why *r* is insufficient. A few responses did not "use the diagram in (i)" to as an example to explain why *r* is insufficient to indicate a linear relationship between two variables. To this candidates are encouraged to read the question carefully.
(iii)	- Using r value to determine	For $D = a + bH^2$, r = 0.87972 = 0.880 (3sf)	This part of Q13 is fairly well-attempted

13	**CORRELATION**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
	which model is a better fit	For $D = a + b\ln H$, r = 0.99898 = 0.999 (3sf) Since the second r value is closer to +1, the second model $D = a + b\ln H$ would provide a better estimate as the transformed points will be more closely to the relevant regression line. Alternative solution : Scatter plot of D against H^2.  Scatter plot of D against $\ln H$.  The transformed data points on a scatter plot of D against $\ln H$	in general, and many of the responses received for this part could correctly identify the second model $D = a + b\ln H$, as the one that would provide a better estimate of D for a given value of H. Some notable mistakes that have been made include : • imprecisely stating that the r-value for the second model being closer to 1 as being indicative of "a better linear relationship" (while a more precise statement could refer to a stronger linear relationship btw. D and $\ln H$, than that btw. D and H^2.) • providing a qualitative explanation for the suitability of the second model, $D = a + b\ln H$, in fitting the data points, but without adequately elaborating why the first model $D = a + bH^2$ would be less suitable for fitting the data in comparison.

13	**CORRELATION**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
		exhibits a stronger / more closely-fitted linear trend, than does the transformed data points on a scatter plot of D against H^2. As such, the second model of a linear relationship between D and $\ln H$, as is expressed by the equation $D = a + b\ln H$, could provide a better estimate of D given a value of H.	
(iv)	- Know how to select appropriate line of regression - To estimate a variable given another	As H is the controlled variable, we should use the regression line of D on $\ln H$: $D = (-394.451...) + (139.049...)\ln H$. When D = 400, $400 = (-394.451...) + (139.049...)\ln H$, Estimated value of H = 302.9197… ≈ 303 (3 s.f.) .	This part of Q13 has been fairly well-answered, with occasional mistakes, and the more notable of which are as follows : • Erroneously using the regression line of D on H instead of the "the model chosen in (iii). • Erroneously choosing to use the regression line of $\ln H$ on D, i.e. $\ln H = (2.8431...) + (0.0071770...)D$, which is less appropriate than using the regression line of D on $\ln H$ since H is a controlled experimental variable in this problem's context. • Erroneously swapping or mixing up the lists on the GC during a "LinReg" calculation, although the model chosen in (iii) and an appropriate line of regression has been correctly selected, as is evident from the usage of a correct form of the regression model / equation (D on $\ln H$), but with mis-calculated regression coefficients whose numerical values could only be

13	CORRELATION		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
			obtained by doing a regression on another regression model (ln H on D), i.e. $D = (2.8431...) + (0.07177...)\ln H$. • Mis-identifying the fitted regression coefficient that is a constant in the model as the coefficient of ln H, and mis-identifying the fitted regression coefficient that is a coefficient of ln H in the model as the constant, i.e. $D = (139.049...) + (-394.451...)\ln H$. • Subjecting the estimate of H to rounding-off error, by giving the regression coefficient to too few decimal places or significant figures. (Due to the nature of the model involving a natural logarithm function, leaving the regression coefficients to 5/6 sig. fig. would be reasonably adequate in producing an accurate estimate for H to 3 s.f., when D = 400). • Erroneously substituted the D and H values of two chosen data points to solve for the coefficients a and b in the model $D = a + b\ln H$, when a least squares regression equation of D on ln H is required.

13	**CORRELATION**		
Part	**Assessment Objectives**	**Mark Scheme**	**Feedback**
(v)	- Know how to comment on reliability of estimates HOT	As evaluating an estimate of H would involve an **extrapolation** of the fitted regression line when $D = 700$, this estimate would be **less reliable** than the estimate of H in (iv), whose evaluation involved an **interpolation**, when $D = 400$. (Mark awarded if student did not state "interpolation". However it must be highlighted to students that they must state for both.)	This part of Q13 has been fairly well-answered in general. Of the set of responses received for this part, a large number redundantly evaluated an estimated value of H when $D = 700$, though this is actually not a requirement of this part of the problem that has requested only for a comment of this estimate's reliability (in comparison with the reliability of H's estimated value when $D = 400$ in part (iv)) – which can be qualitatively determined as less reliable if the estimate of H is an extrapolated value or determined as more reliable of the estimate of H is an interpolated value. Careful reading of the requirements specified in this part of the question would have been helpful in saving some time that was spent calculating the above-mentioned estimate of H. Some of the received responses happened to be penalised for not having adequately made a *comparison*, especially by merely explaining how the estimate of H when $D = 700$ is an unreliable one (due to this estimate being an extrapolated value) but without putting down any mention about the estimate of H when $D = 400$ being an interpolated value (that is consequently more reliable) in contrast.

CATHOLIC JUNIOR COLLEGE
H2 MATHEMATICS 9740
JC2 PRELIM PAPER 2 SOLUTIONS 2012