2024 GCE A-Level H2 Mathematics 9758 Paper 2

Section A: Pure Mathematics [40 marks]

1 **Do not use a calculator in answering this question.**

The complex number $-2+\mathrm{i}$ is denoted by z.

(a) Find the real numbers a and b such that $z = az^* + b$. [2]

Solutions
$z=-2+\mathrm{i}$ $z=az^*+b$ $-2+\mathrm{i}=a(-2-\mathrm{i})+b$ $\quad=-2a+b-a\mathrm{i}$ Comparing imaginary and real parts respectively: $a=\underline{-1}$ $-2=-2a+b \Rightarrow b=-2+2(-1)=\underline{-4}$

The complex number $1-3\mathrm{i}$ is denoted by w.

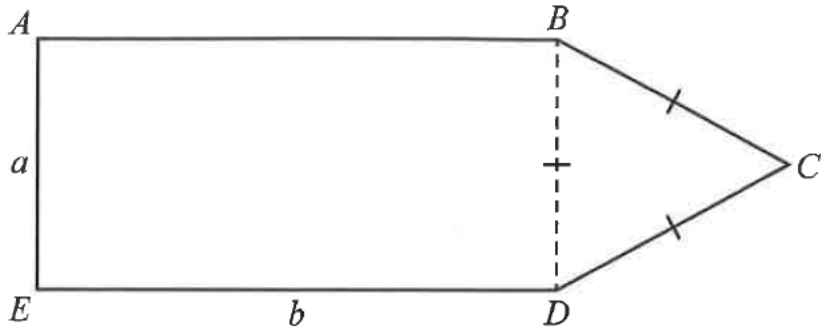
(b) Without using a calculator, evaluate $wz-\dfrac{w}{z}$. Give your answer in the form $c+d\mathrm{i}$, where c and d are real numbers. [4]

Solutions

$$\begin{aligned} wz-\frac{w}{z} &= (1-3\mathrm{i})(-2+\mathrm{i})-\frac{1-3\mathrm{i}}{-2+\mathrm{i}} \\ &= -2+\mathrm{i}+6\mathrm{i}+3-\left(\frac{1-3\mathrm{i}}{-2+\mathrm{i}}\times\frac{-2-\mathrm{i}}{-2-\mathrm{i}}\right) \\ &= 1+7\mathrm{i}-\frac{-2-\mathrm{i}+6\mathrm{i}-3}{4+1} \\ &= 1+7\mathrm{i}-\frac{-5+5\mathrm{i}}{5} \\ &= 1+7\mathrm{i}+1-\mathrm{i} \\ &= \underline{2+6\mathrm{i}} \end{aligned}$$

(i.e. $c=2$, $d=6$)

2



A gardener designs a flower bed *ABCDE* in the shape of a rectangle with an equilateral triangle on one of the shorter sides. Side *AE* is of length a m and side *ED* is of length b m (see diagram). The total perimeter of the flower bed is 20 m.

Find the maximum possible area of the flower bed, showing that it is a maximum value. Give your answer correct to 4 significant figures. [6]

Solutions

Perimeter $=3a+2b=20 \Rightarrow b=10-\dfrac{3a}{2}$

Let area of flower bed be y m^2.

$$y=ab+\frac{1}{2}(a)(a)\sin 60^\circ$$

$$=a\left(10-\frac{3a}{2}\right)+\frac{a^2}{2}\left(\frac{\sqrt{3}}{2}\right)$$

$$=10a+\frac{\left(\sqrt{3}-6\right)a^2}{4}$$

$$\frac{\mathrm{d}y}{\mathrm{d}a}=10+\frac{\left(\sqrt{3}-6\right)a}{2}$$

When y is maximum, $\dfrac{\mathrm{d}y}{\mathrm{d}a}=0$

$$10+\frac{\left(\sqrt{3}-6\right)a}{2}=0$$

$$\frac{\left(6-\sqrt{3}\right)a}{2}=10$$

$$a=\frac{20}{6-\sqrt{3}}$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}a^2}=\frac{\sqrt{3}-6}{2}\approx -2.13<0$$

$\Rightarrow$ The area y m^2 is a maximum when $a=\dfrac{20}{6-\sqrt{3}}$.

$$\text{Maximum area}=10\left(\frac{20}{6-\sqrt{3}}\right)+\frac{\left(\sqrt{3}-6\right)}{4}\left(\frac{20}{6-\sqrt{3}}\right)^2$$

$$=\underline{23.43\text{ m}^2}\ \text{(4 s.f.)}$$

Solutions
Comment: From what was computed above, $a = 4.68$, $b = 2.97$ (3 s.f.)   But question says that the equilateral triangle is on the shorter side. In fact, based on how the question is phrased, there is no maximum value for the area.
Non-Calculus Method $y = 10a + \dfrac{(\sqrt{3}-6)a^2}{4}$ Let $k = \dfrac{6-\sqrt{3}}{4} (\approx 1.067)$ $y = 10a - ka^2$ $= -k\left(a^2 - \dfrac{10}{k}a\right)$ $= -k\left(a - \dfrac{5}{k}\right)^2 + k\left(\dfrac{5}{k}\right)^2$ $= -k\left(a - \dfrac{5}{k}\right)^2 + \dfrac{25}{k}$ The area y is a quadratic expression in a with a negative coefficient for a^2. Hence it has a maximum turning point at $\left(\dfrac{5}{k}, \dfrac{25}{k}\right)$.   Hence maximum possible area $= 25\left(\dfrac{4}{6-\sqrt{3}}\right)$ $= \underline{23.43 \text{ m}^2}$ (to 4 s.f.)

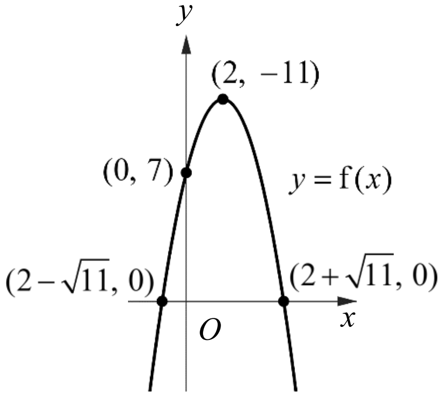
3 The function f is such that $\mathrm{f}(x) = -x^2 + 4x + 7$, for $x \in \mathbb{R}$.

(a) Find the range of f. [2]

Solutions
$\mathrm{f}(x) = -x^2 + 4x + 7$ $= -(x-2)^2 + 4 + 7$ $= -(x-2)^2 + 11$ ≤ 11 $\therefore$ Range of f is $R_\mathrm{f} = \underline{(-\infty, 11]}$

(b) Sketch the graph of f, giving the exact coordinates of the points where the curve crosses the axes. [2]

Solutions



$y = \mathrm{f}(x) = -x^2 + 4x + 7$

Let $x = 0$: $y = 7 \Rightarrow y$-intercept $= \underline{(0,\ 7)}$

Let $y = 0$: $-(x-2)^2 + 11 = 0$

$$x - 2 = \pm\sqrt{11}$$

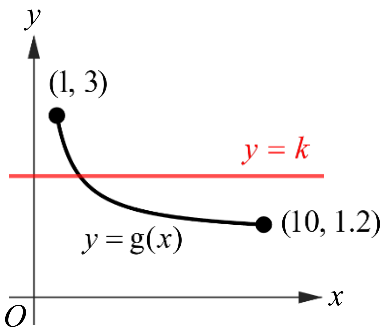
$$x = 2 \pm \sqrt{11}$$

$\Rightarrow x$-intercepts $= \underline{(2-\sqrt{11},\ 0)}$ and $\underline{(2+\sqrt{11},\ 0)}$

The function g is such that $\mathrm{g}(x) = \dfrac{2}{x} + 1$, for $1 \le x \le 10$.

(c) Explain how you know g^{-1} exists. [1]

Solutions



Any horizontal line $y = k$, where $k \in \mathbb{R}$, cuts the graph of $y = \mathrm{g}(x)$ at most once. Hence g is a one-to-one function, and hence g^{-1} exists.

<u>Alternative</u>

$\mathrm{g}'(x) = -\dfrac{2}{x^2} < 0$ for $1 \le x \le 10$

This shows g is a decreasing function, hence it is a one-to-one function and g^{-1} exists.

(d) Find the value of $\mathrm{fg}^{-1}(1.5)$. [3]

Solutions
Let $y = \mathrm{g}(x) = \dfrac{2}{x} + 1$ $\Rightarrow x = \dfrac{2}{y-1} = \mathrm{g}^{-1}(y)$ $\Rightarrow \mathrm{g}^{-1}(x) = \dfrac{2}{x-1}$ for $1.2 \le x \le 3$ $\mathrm{fg}^{-1}(1.5) = \mathrm{f}\left(\dfrac{2}{1.5-1}\right)$ $= \mathrm{f}(4)$ $= -(4-2)^2 + 11$ $= \underline{7}$

4 The sum of the first n terms of the series T is given by $2n^3 - 8n^2 - 4n$.

(a) Find an expression for t_n, the nth term of series T. [2]

Solutions
$t_n = 2n^3 - 8n^2 - 4n - \left(2(n-1)^3 - 8(n-1)^2 - 4(n-1)\right)$ $= 2n^3 - 8n^2 - 4n - 2(n^3 - 3n^2 + 3n - 1)$ $+8(n^2 - 2n + 1) + 4n - 4$ $= -2(-3n^2 + 3n - 1) + 8\left(-2n + 1\right) - 4$ $= 6n^2 - 6n + 2 - 16n + 8 - 4$ $= \underline{6n^2 - 22n + 6}$ (Check: When $n = 1$, $2(1)^3 - 8(1)^2 - 4(1) = -10$ and $6(1) - 22(1) + 6 = -10$ $\therefore$ Expression for t_n holds for $n \ge 1$)

The nth term of the series U is given by $u_n = 50n - 204$.

(b) Find the values of n for which $u_n = t_n$. [2]

Solutions
$u_n = t_n$ $50n - 204 = 6n^2 - 22n + 6$ $6n^2 - 72n + 210 = 0$ $6(n-5)(n-7) = 0$ $\therefore\ n = \underline{5}$ or $\underline{7}$

4 [Continued]

The nth term of the series V is given by $v_n = 3n + 16$.

(c) Find the smallest number greater than 100 that is in both series U and series V. [2]

Solutions
Method 1 $u_n = 50n - 204 > 100 \Rightarrow n > \frac{304}{50} = 6.08$ $\Rightarrow u_n > 100$ for all $n \geq 7$ $v_n = 3n + 16 > 100 \Rightarrow n > \frac{84}{3} = 28$ $\Rightarrow v_n > 100$ for all $n \geq 29$ Using GC, the smallest number greater than 100 that is in both series U and series V is 196 ($u_8 = v_{60} = 196$). Method 2 Series U is an arithmetic series with common difference 50. Series V is an arithmetic series with common difference 3. From GC, $u_5 = v_{10} = 46$ is the smallest number that exists in both series U and series V. To get the next number that exists in both series, we add the lowest common multiple of 50 and 3, i.e. 150. $\therefore\ 46 + 150 = 196$ is the smallest number greater than 100 that exists in both series.

The nth term of the series W is given by $w_n = 3n^2 - 5n + 7$.

(d) **(i)** Explain why all the terms in series W are odd numbers. [2]

Solutions
$w_n = 3n^2 - 5n + 7$ When n is odd, $w_n = \underbrace{(3n^2 + 7)}_{\text{even}} - \underbrace{5n}_{\text{odd}}$ is odd. When n is even, $w_n = \underbrace{(3n^2 + 7)}_{\text{odd}} - \underbrace{5n}_{\text{even}}$ is odd. Hence all terms in series W are odd numbers.

Solutions
Alternative Method 1 Case 1: n is even (i.e. $n=2k$ where $k\in\mathbb{Z}^+$) $w_{2k}=3(2k)^2-5(2k)+7$ $=12k^2-10k+7$ $=2(6k^2-5k+3)+1$ $=2K_k+1$ where $K_k=6k^2-5k+3$ is an integer. Hence w_{2k} is an odd number. Case 2: n is odd (i.e. $n=2k-1$ where $k\in\mathbb{Z}^+$) $w_{2k-1}=3(2k-1)^2-5(2k-1)+7$ $=12k^2-12k+3-10k+5+7$ $=12k^2-22k+15$ $=2(6k^2-11k+7)+1$ $=2L_k+1$ where $L_k=6k^2-11k+7$ is an integer. Hence w_{2k-1} is an odd number. Hence all terms in series W are odd numbers. Alternative Method 2 *(induction argument)* $w_n-w_{n-1}=3n^2-5n+7-3(n-1)^2+5(n-1)-7$ $=6n-3-5$ $=6n-8$ $=2(3n-4)$ $\Rightarrow$ The difference of any two consecutive terms in W is even. $w_1=3(1)^2-5(1)+7=5$, which is odd. Since an odd and even number add up to produce an odd number, (by induction) all the terms of W are odd.

(ii) Hence explain why series U and series W do not have any terms in common. [1]

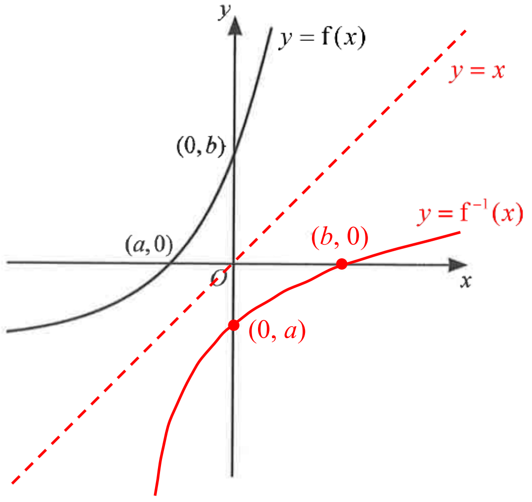
Solutions
$u_n=50n-204$ $=2(25n-102)$ $=2M_n$ where $M_n=25n-102$ is an integer. Hence all terms in series U are even numbers. Since all terms in series W are odd numbers, and a number cannot be both even and odd, there are no common terms in series U and series W.

5 **(a)** The graph of $y = \mathrm{f}(x)$ intersects the x-axis at the point $(a,\ 0)$ and the y-axis at the point $(0,\ b)$.

(i) The graph of $y = \mathrm{f}(x)$ is shown in Fig. 1. The scales on the x- and y-axes are the same.

Sketch the graph of $y = \mathrm{f}^{-1}(x)$ on Fig. 1, labelling the intersection with the axes. [1]

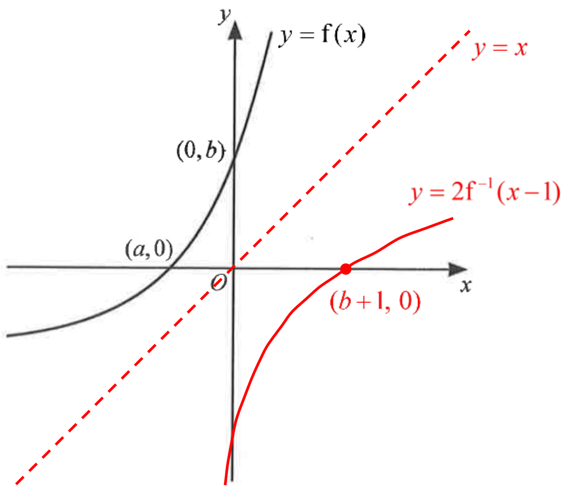
Solutions



(ii) The graph of $y = \mathrm{f}(x)$ is also shown in Fig. 2. The scales on the x- and y-axes are the same.

Sketch the graph of $y = 2\mathrm{f}^{-1}(x-1)$ on Fig. 2, labelling the intersection with the x-axis. [2]

Solutions

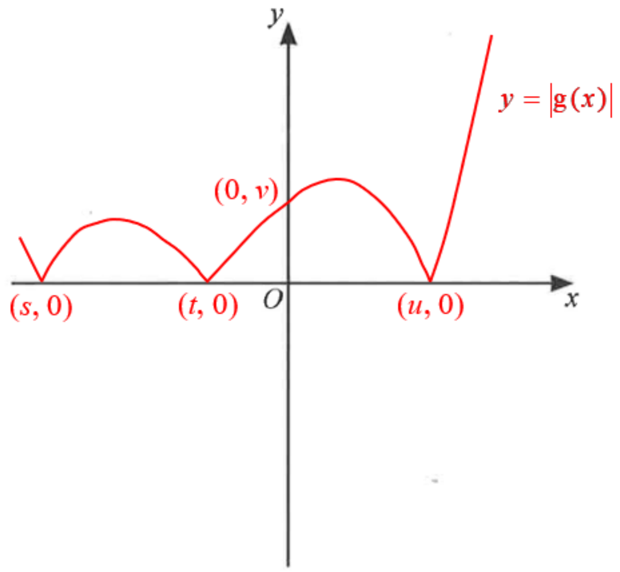


5 **(b)** The graph of $y = \mathrm{g}(x)$ intersects the x-axis at the points $(s,\ 0)$, $(t,\ 0)$ and $(u,\ 0)$ and the y-axis at the point $(0,\ v)$.

(i) The graph of $y = \mathrm{g}(x)$ is shown in Fig. 3.1.

Sketch the graph of $y = |\mathrm{g}(x)|$ on Fig. 3.2, labelling the points where the graph intersects or touches the axes. [2]

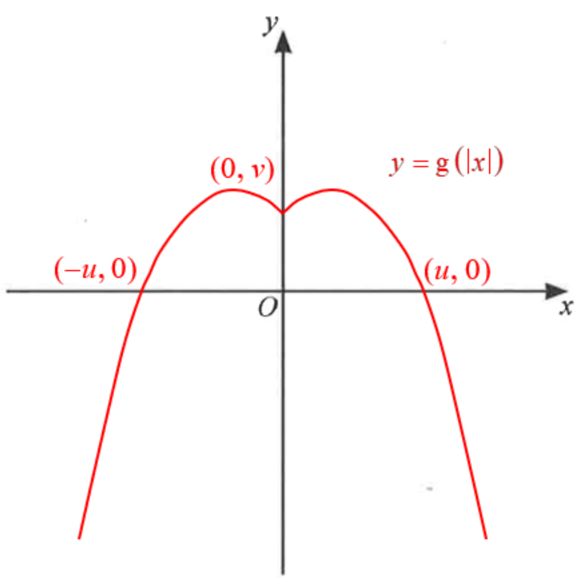
Solutions



(ii) The graph of $y = \mathrm{g}(x)$ is also shown in Fig. 4.1.

Sketch the graph of $y = \mathrm{g}(|x|)$ on Fig. 4.2, labelling the points where the graph intersects or touches the axes. [2]

Solutions

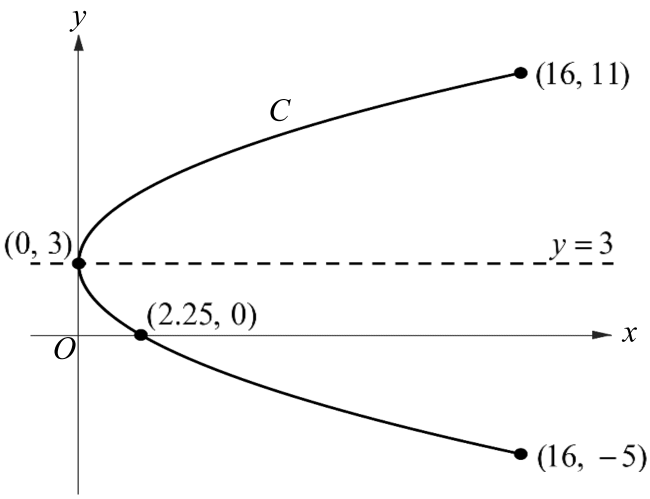


5 **(c)** The curve C has parametric equations

$$x = t^2, \qquad y = 2t + 3, \qquad \text{for } -4 \le t \le 4.$$

(i) Sketch the graph of C. [1]

(ii) Mark on your sketch the coordinates of the points where the curve C meets the axes. [2]

Solutions
 Let $y = 0$: $2t + 3 = 0 \Rightarrow t = -1.5$ $\Rightarrow x = (-1.5)^2 = 2.25$ $\therefore$ x-intercept $= \underline{(2.25,\ 0)}$ Let $x = 0$: $t = 0 \Rightarrow y = 3$ $\therefore$ y-intercept $= \underline{(0,\ 3)}$

(iii) State the equations of any lines of symmetry. [1]

Solutions
$\underline{y = 3}$

Section B: Probability and Statistics [60 marks]

6 A sports club has three categories of membership. There are 27 Youth members, 45 Adult members and 18 Senior members.

The club secretary wishes to find out the opinions of members of the club about the facilities offered. She gives a questionnaire to all the members of the club. She receives replies from 65 of the members.

(a) Explain whether the 65 members comprise a sample or a population. [1]

Solutions
The 65 members comprise a sample as it does not cover all 90 members of the club (the population for the secretary's survey). OR The 65 members comprise a sample as it is a proper subset of the 90 members of the club (the population for the secretary's survey).

The secretary decides to form a committee of members to discuss the results of her questionnaire.

(b) Explain an advantage of choosing a random sample of members for her committee. [1]

Solutions
Having a random sample minimises the possibility of bias, since members are independently chosen and each member has an equal probability of being selected for the committee.

The secretary decides that the committee will consist of 6 randomly chosen members, but there will be:

- at least 1 member from each category
- more Adult members than Youth members and
- more Adult members than Senior members.

(c) Find the probability that the committee contains more Youth members than Senior members. [4]

Solutions
Possible make-up of committees: YAAAAS, YAAASS, YYAAAS No. of YAAAAS committees = ${}^{27}C_1 \times {}^{45}C_4 \times {}^{18}C_1$ No. of YAAASS committees $= {}^{27}C_1 \times {}^{45}C_3 \times {}^{18}C_2$ No. of YYAAAS committees $= {}^{27}C_2 \times {}^{45}C_3 \times {}^{18}C_1$

Probability that committee contains more Youth than Senior members

$$= \frac{{}^{27}C_2 \times {}^{45}C_3 \times {}^{18}C_1}{({}^{27}C_1 \times {}^{45}C_4 \times {}^{18}C_1) + ({}^{27}C_1 \times {}^{45}C_3 \times {}^{18}C_2) + ({}^{27}C_2 \times {}^{45}C_3 \times {}^{18}C_1)}$$

$= \underline{0.40625}$ or $\underline{\dfrac{13}{32}}$

7 **(a)** A, B and C are taking part in a game in which a prize is hidden in one of n identical closed boxes, where $n > 4$. First, A opens at random two boxes in an attempt to find the prize. If A fails to find the prize, B then opens at random two of the remaining boxes in an attempt to find the prize. If B fails to find the prize, C opens all the remaining boxes and finds the prize.

(i) Find, in terms of n, the probability that A finds the prize. [1]

Solutions
P(A finds the prize) $=$ P(A finds prize in 1st box opened) +P(A finds prize in 2nd box opened) $= \dfrac{1}{n} + \left(\dfrac{n-1}{n}\right)\left(\dfrac{1}{n-1}\right)$ $= \underline{\dfrac{2}{n}}$

(ii) Show that the probability that B finds the prize is $\dfrac{2}{n}$. [1]

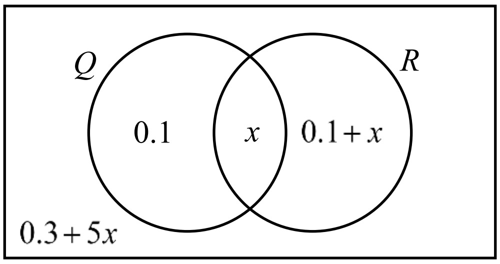
Solutions
P(A does not find the prize) $= 1 - \dfrac{2}{n}$ $= \dfrac{n-2}{n}$ P(B finds the prize) $=$ P(A loses, B finds prize in 1st box opened) +P(A loses, B finds prize in 2nd box opened) $= \left(\dfrac{n-2}{n}\right)\left(\dfrac{1}{n-2}\right) + \left(\dfrac{n-2}{n}\right)\left(\dfrac{n-3}{n-2}\right)\left(\dfrac{1}{n-3}\right)$ $= \dfrac{1}{n} + \dfrac{1}{n}$ $= \underline{\dfrac{2}{n}}$ (shown)

(iii) Find the range of values of n for which the probability that C finds the prize is more than 8 times the probability that either A or B find the prize. [3]

Solutions
P(C finds the prize) $=1-\text{P(A finds the prize)}-\text{P(B finds the prize)}$ $=1-\dfrac{2}{n}-\dfrac{2}{n}$ $=1-\dfrac{4}{n}$ P(C finds the prize) > 8 × P(either A or B finds the prize) $1-\dfrac{4}{n}>8\left(\dfrac{2}{n}+\dfrac{2}{n}\right)$ $1>\dfrac{36}{n}$ $n>36$ $\therefore$ Set of values of n $=\underline{\{n\in\mathbb{Z}:n\geq 37\}}$

(b) The events Q and R are such that $\text{P}(Q\cap R)=x$, $\text{P}(Q\cap R')=0.1$, $\text{P}(Q'\cap R)=\text{P}(Q)$ and $\text{P}((Q\cup R)')=\text{P}(Q)+2\text{P}(R)$. By using a Venn diagram, or otherwise, find the exact value of x. [3]

Solutions



$\text{P}(Q\cap R)=x$

$\text{P}(Q\cap R')=0.1$

$\text{P}(Q'\cap R)=\text{P}(Q)$

$\qquad\qquad=0.1+x$

$\text{P}((Q\cup R)')=\text{P}(Q)+2\text{P}(R)$

$\qquad\qquad=0.1+x+2(x+0.1+x)$

$\qquad\qquad=0.3+5x$

Total probability = 1

$0.1+x+0.1+x+0.3+5x=1$

$7x=0.5=\dfrac{1}{2}$

$x=\underline{\dfrac{1}{14}}$

8 Lee has a bird feeder on his balcony. Every morning he spends 20 minutes, while he eats his breakfast, counting the number of birds that visit the feeder. Over many months he has found that the mean number of birds visiting the feeder in this time interval is 17.3. When the bird feeder was damaged in a storm, Lee replaced it with a new bird feeder.

He suspects that the mean number of birds visiting the new bird feeder while he eats his breakfast has reduced. He decides to check this with a hypothesis test at the $\alpha\%$ level of significance, where α is an integer. He records the number of birds, x, visiting the feeder in 20 minutes each morning for a random sample of n mornings.

(a) Explain whether Lee should carry out a one-tailed test or a two-tailed test. [1]

Solutions
Lee should carry out a <u>one-tailed test</u> as he wants to test if the mean number of birds visiting the new bird feeder has <u>reduced</u>, i.e. less than the current mean number of 17.3.

(b) State hypotheses for Lee's test, defining any parameters that you use. [2]

Solutions
Let X be the random variable of the number of birds visiting Lee's bird feeder during a 20-minute period in a morning. Test $\mathrm{H_0}: \mu = 17.3$ against $\mathrm{H_1}: \mu < 17.3$ at $\alpha\%$ level of significance. μ is the population mean number of birds that visit Lee's bird feeder during a 20-minute period in the morning.

Here is a summary of the data Lee collected.

$$n = 32 \qquad \Sigma x = 512 \qquad \Sigma x^2 = 8702$$

Lee carried out his test and concluded that the null hypothesis should be rejected.

(c) Calculate unbiased estimates of the population mean and variance of the number of birds, and determine the minimum possible value of the integer α. [3]

Solutions
Unbiased estimate of population mean, $\bar{x} = \frac{512}{32} = \underline{16}$ Unbiased estimate of population variance, $s^2 = \frac{1}{31}\left(8702 - \frac{512^2}{32}\right) = \underline{\frac{510}{31}}$ or $\underline{16.5}$ (3 s.f.) Under $\mathrm{H_0}$, since the sample size of 32 is large enough, $\bar{X} \sim \mathrm{N}\left(17.3, \frac{16.4516129}{32}\right)$ approximately by Central Limit Theorem.

p-value $= \mathrm{P}(\bar{X} < 16) = 0.0349$ (3 s.f.)
Since H_0 is rejected, p-value < $\dfrac{\alpha}{100}$ $\Rightarrow \alpha > 3.49$ $\therefore$ Smallest possible value of $\alpha = \underline{4}$ [Standardised test statistic $Z = \dfrac{\bar{X} - 17.3}{\sqrt{\dfrac{16.4516129}{32}}} \sim \mathrm{N}(0, 1)$ $z = \dfrac{16 - 17.3}{\sqrt{\dfrac{16.4516129}{32}}} = -1.813$]

(d) State the conclusion to Lee's test in the context of the question. [1]

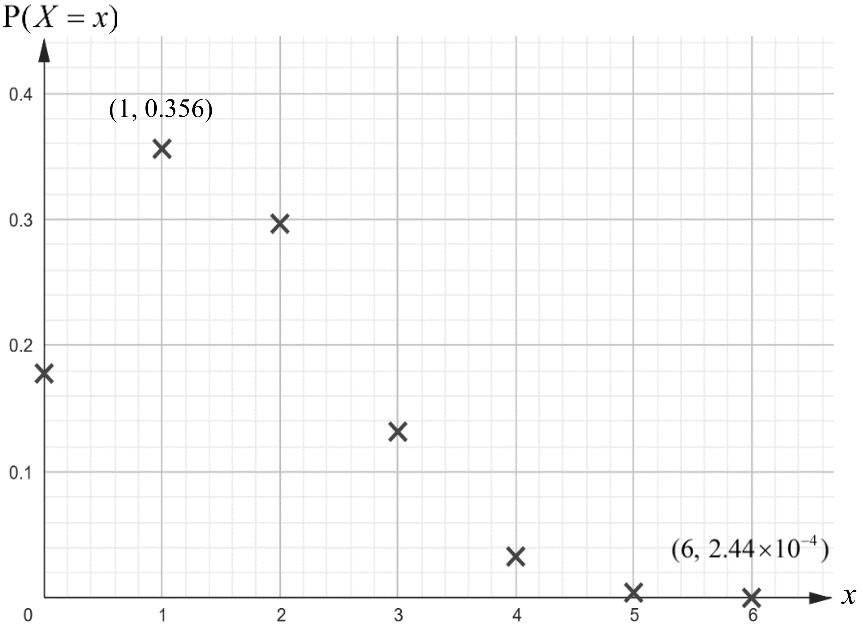
Solutions
There is sufficient evidence at α% level of significance to conclude that the mean number of birds visiting the new bird feeder while Lee eats his breakfast has reduced.

(e) Explain whether using the number of birds for each of 10 mornings would have been sufficient for Lee to carry out his hypothesis test. [1]

Solutions
Data from 10 mornings is insufficient to carry out the hypothesis test. As the population distribution of X is unknown, a sample size of 10 is not large enough for Central Limit Theorem to apply so that $\bar{X}$ is approximately normally distributed.

9 **(a)** The random variable X has the distribution $\text{B}(6,\ 0.25)$. Sketch the distribution of X. [2]

Solutions



(b) One of the jobs of a quality control operative in a food-processing factory is to check the quality of a random sample of 50 meat pies from the production line each day. She records the number of pies found to be 'unsatisfactory' each day for 80 days. Her results are shown in the table below.

Number unsatisfactory	0	1	2	3	4	5	6 or more
Frequency	57	9	6	4	3	1	0

Use the information in the table to estimate the probability that a randomly chosen pie is unsatisfactory. [2]

Solutions

Estimated probability that a randomly chosen pie is unsatisfactory

$$=\frac{(0\times 57)+(1\times 9)+(2\times 6)+(3\times 4)+(4\times 3)+(5\times 1)}{50\times 80}$$

$$=\underline{\frac{1}{80}} \text{ or } \underline{0.0125}$$

(c) One of the products made in the food-processing factory is the 'Frozen Cheesee Burger'. A fixed number of the burgers are tested each day and the number found to have insufficient cheese in them is denoted by Y.

(i) State, in context, two assumptions needed for Y to be well modelled by a binomial distribution. [2]

Solutions
1. Whether a burger has insufficient cheese in it is independent of whether any of the other burgers has insufficient cheese in them. 2. The probability of a randomly chosen burger having insufficient cheese is a constant.

Assume now that Y has the distribution $\mathrm{B}(120,\ 0.03)$.

(ii) Find the probability that, on a randomly chosen day, fewer than 3 burgers are found to have insufficient cheese in them. [1]

Solutions

$$\begin{aligned}\mathrm{P}(Y<3) &= \mathrm{P}(Y\le 2)\\ &= 0.298435\\ &= \underline{0.298}\ \text{(3 s.f.)}\end{aligned}$$

(iii) For a randomly chosen period of 28 days, find the expectation of the number of days on which fewer than 3 burgers have insufficient cheese in them. [1]

Solutions

Expected number of days (out of 28 days) with fewer than 3 burgers with insufficient cheese

$= 28\times \mathrm{P}(Y\le 2)$

$= \underline{8.36}$ (3 s.f.)

(iv) Find the probability that, in a randomly chosen period of 28 days, more than 100 burgers are found to have insufficient cheese in them. [2]

Solutions

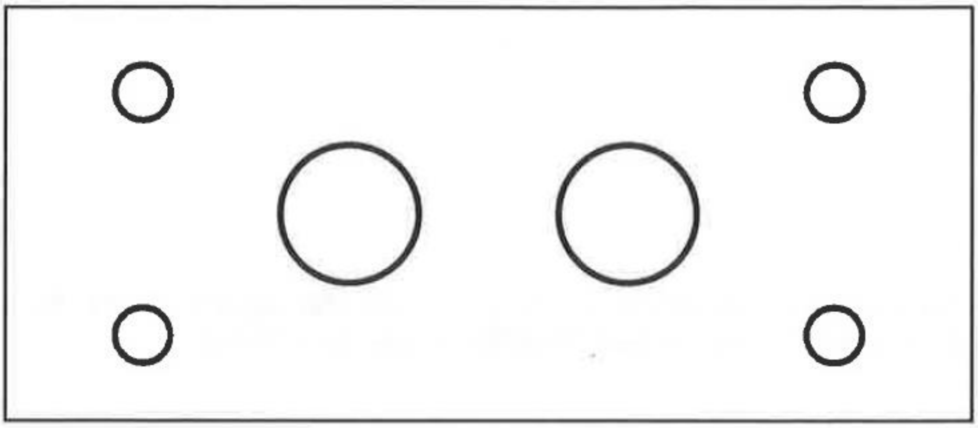
Let W denote the random variable of the number of burgers (out of $28\times 120 = 3360$ burgers) with insufficient cheese in them.

$W \sim \mathrm{B}(3360,\ 0.03)$

$$\begin{aligned}\mathrm{P}(W>100) &= 1-\mathrm{P}(W\le 100)\\ &= \underline{0.506}\ \text{(3 s.f.)}\end{aligned}$$

10 In this question you should state the parameters of any distributions you use.

A company makes metal plates that can be used to fix fence panels onto posts. The metal plates have four small holes drilled into them for screws and two large holes drilled into them for bolts (see diagram).



Before the holes are drilled, the masses of plates, in grams, follow the distribution $N(200, 1.6^2)$.

(a) Find the probability that the mass of a randomly chosen plate before drilling is more than 197.5 grams. [1]

Solutions
Let X denote the random variable of mass (in grams) of a randomly chosen plate before drilling. $X \sim \underline{N(200, 1.6^2)}$ $P(X > 197.5) = \underline{0.941}$ (3 s.f.)

Drilling the holes reduces the mass of each plate by 5%. A production worker selects 8 of the drilled plates at random.

(b) Find the probability that at least 5 of these 8 plates have masses between 190 grams and 192 grams. [4]

Solutions
Mass of a randomly chosen plate after drilling $= 0.95X$ $0.95X \sim N(0.95 \times 200, 0.95^2 \times 1.6^2) = \underline{N(190, 1.52^2)}$ Let Y denote the random variable of number of plates (out of 8 plates) with masses between 190 grams and 192 grams after drilling. $Y \sim B(8, P(190 < 0.95X < 192)) = \underline{B(8, 0.405878)}$ $P(Y \geq 5) = 1 - P(Y \leq 4)$ $= \underline{0.183}$ (3 s.f.)

The drilled plates are sold in packs of 20 randomly chosen plates.

(c) Find the probability that the total mass of a pack of 20 drilled plates is less than 3805 grams. [3]

Solutions
Total mass of pack of 20 drilled plates $= T = 0.95X_1 + 0.95X_2 + ... + 0.95X_{20}$ $\text{E}(T) = 20 \times 190 = 3800$ $\text{Var}(T) = 20 \times 1.52^2 = 46.208$ $\Rightarrow T \sim \underline{\text{N}(3800,\ 46.208)}$ $\text{P}(T < 3805) = \underline{0.769}$ (3 s.f.)

The manufacturer decides to sell 'Value' packs containing all the materials needed to fix fence panels onto posts. Each Value pack consists of 20 drilled plates together with the right number of screws and bolts to fit them; all of these are randomly chosen.

- The masses of screws, in grams, follow the distribution $\text{N}(10,\ 0.3^2)$.
- The masses of bolts, in grams, follow the distribution $\text{N}(44,\ 1.1^2)$.

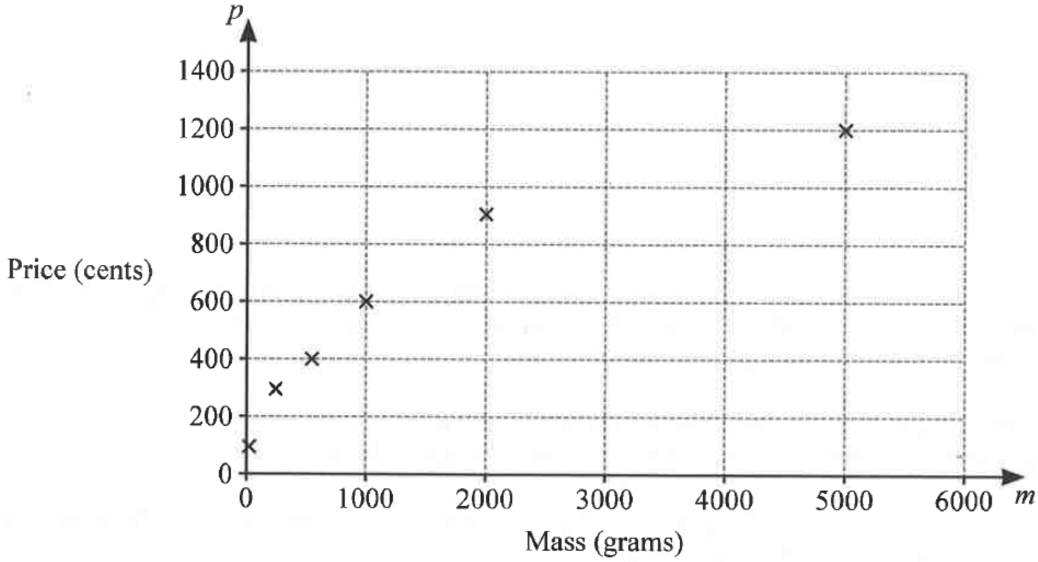
(d) Find the mass exceeded by just 5% of the Value packs. Give your answer to the nearest gram. You should ignore the mass of any packaging. [4]

Solutions
Let V denote the random variable of the mass of a randomly chosen Value pack. $\text{E}(V) = 3800 + (4 \times 20 \times 10) + (2 \times 20 \times 44)$ $= 6360$ $\text{Var}(V) = 46.208 + (4 \times 20 \times 0.3^2) + (2 \times 20 \times 1.1^2)$ $= 101.808$ $\Rightarrow V \sim \underline{\text{N}(6360,\ 101.808)}$ $\text{P}(V > v) = 0.05$ Using invNorm on GC, $v = 6376.596565$ $= \underline{6377\text{ g}}$ (nearest gram)

11 **(a)** A certain brand of breakfast cereal is sold in a variety of different sized packs. Details of the mass of cereal in each pack, m grams, and the price, p cents, are given in the table below.

m	24	250	550	1000	2000	5000
p	100	300	400	600	900	1200

A scatter diagram for the data is shown below.



(i) Explain what the scatter diagram tells you about the relationship between m and p. [1]

Solutions
The scatter diagram tells us that as mass of a pack of cereal (m) increases, the price (p) increases at a decreasing rate.

The following two models are proposed, where a, b, d and e are constants.

$$p = a + b\ln m \qquad\qquad p = d + e\sqrt{m}$$

(ii) Determine which of these models gives the better fit to the data. State the values of the constants and the product moment correlation coefficient in this case. [4]

Solutions
For $p = a + b\ln m$, product moment correlation coefficient $r_1 = 0.928513$ For $p = d + e\sqrt{m}$, product moment correlation coefficient $r_2 = 0.991438$ Since r_2 is closer to 1 than r_1, the model $\underline{p = d + e\sqrt{m}}$ gives a better fit to the data. For this model, $d = 32.952567 = \underline{33.0}$ and $e = 17.269804 = \underline{17.3}$ $r = r_2 = \underline{0.991}$ (all answers to 3 s.f.)

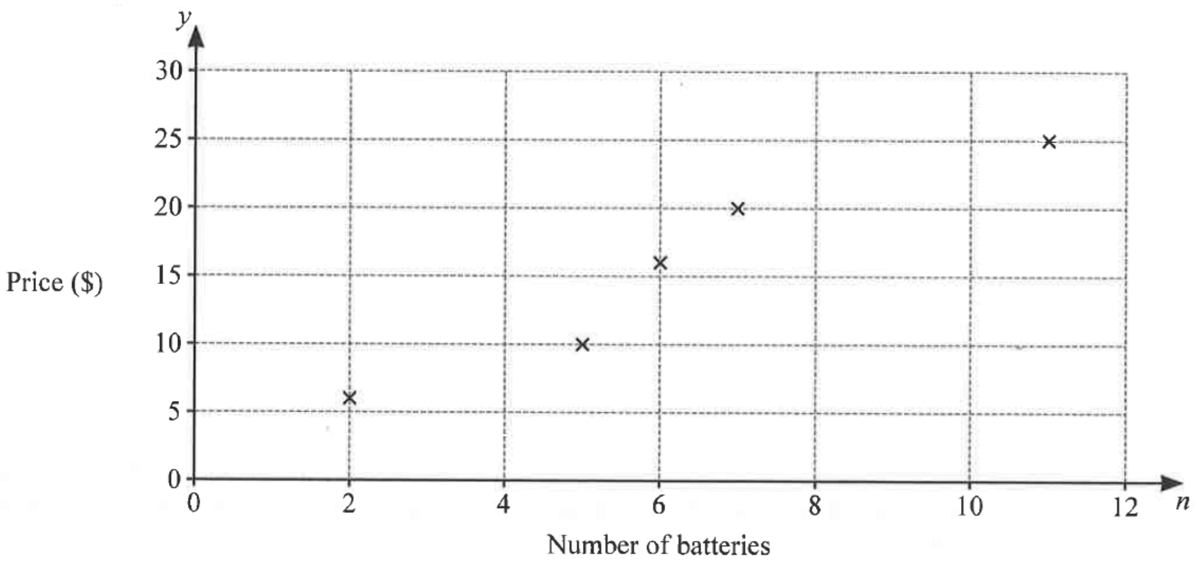
(iii) A new pack containing 750 grams of cereal is introduced. Use the model you identified in part **(a)(ii)** to estimate the price of this pack, correct to the nearest 10 cents. Explain whether your answer is reliable. [2]

Solutions
$p = 32.952567 + 17.269804\sqrt{m}$ Sub. $m = 750$: $p = 32.952567 + 17.269804\sqrt{750}$ $= 505.906$ $\therefore$ Estimated price $= \underline{510 \text{ cents}}$ (nearest 10 cents) This estimate is reliable as the product moment correlation coefficient of 0.991 is very close to 1, indicating a strong positive linear correlation between $\sqrt{m}$ and p. Furthermore, the value $m = 750$ is within the data range $24 \le m \le 5000$, hence the estimate is an interpolation.

(b) Kai is investigating the relationship between the number, n, of a particular type of high-performance batteries in a pack and the price, \$$y$, of the pack she found in different stores. Her results are shown in the table below.

n	2	5	6	7	11
y	6	10	16	20	25

This information is shown in the scatter diagram below.



11 [Continued]

Kai decides to investigate whether the model $y=\frac{5}{2}n+1$ is a good fit for this data.

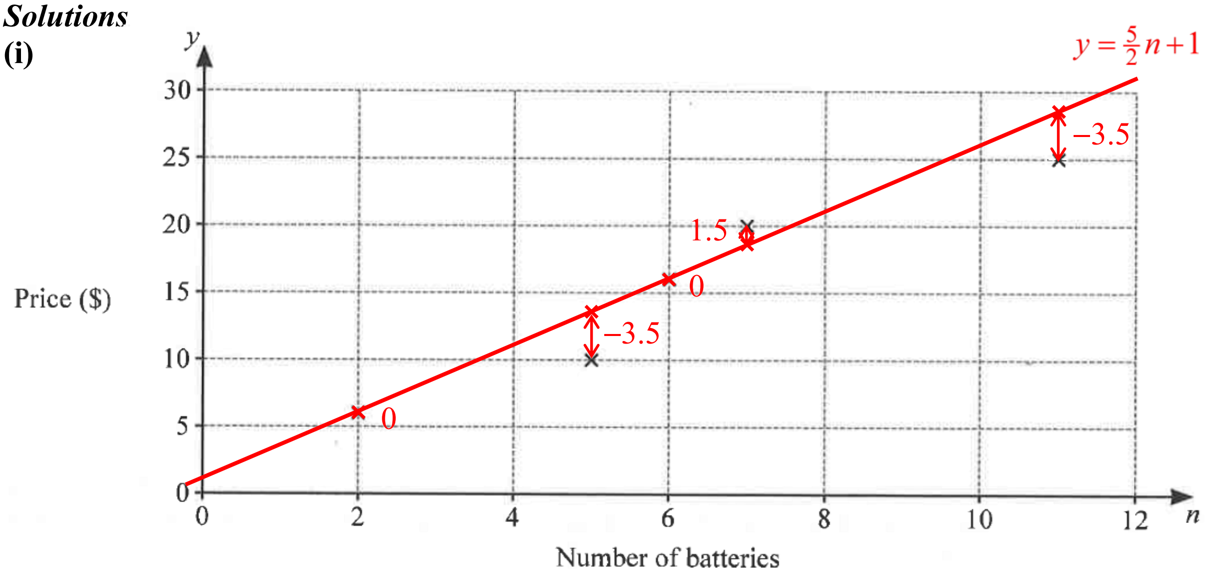
(i) Draw the line $y=\frac{5}{2}n+1$ on the scatter diagram above. [1]

(ii) For the model $y=\mathrm{f}(n)$, the residual for a point (a, b) is $b-\mathrm{f}(a)$.

(A) Mark the residuals for the points on the scatter diagram above. [1]

Solutions

(i)



(ii)(A)

Point (a, b)	(2, 6)	(5, 10)	(6, 16)	(7, 20)	(11, 25)
$\mathrm{f}(a)$	6	13.5	16	18.5	28.5
Residual $b-\mathrm{f}(a)$	0	−3.5	0	1.5	−3.5

(B) Explain why Kai should use the sum of the squares of the residuals rather than the sum of the residuals when assessing the fit of the model. [1]

Solutions

As residuals may be positive or negative (or zero), they may cancel out if Kai simply sums the residuals. Squaring the residuals allows Kai to sum up positive numbers for a more accurate assessment of the fit of the model.

(C) Calculate the sum of the squares of the residuals for the line $y=\frac{5}{2}n+1$. [1]

Solutions
Sum of squares of residuals for the line $y=\frac{5}{2}n+1$ is $0^2+(-3.5)^2+0^2+1.5^2+(-3.5)^2=\underline{26.75}$

Kai's friend points out that other lines parallel to Kai's line can be drawn which are a better fit for the data.

(iii) Explain how the sum of the squares of the residuals for a line that is a better fit for the data differs from the sum of the squares of the residuals for the line $y=\frac{5}{2}n+1$ found in part **(ii)(C)**. [1]

Solutions
The sum of squares of residuals for a line that is a better fit will be less than the sum of squares of residuals for the line $y=\frac{5}{2}n+1$, i.e. less than 26.75.

(iv) Find the range of values of c for which the line $y=\frac{5}{2}n+c$ is a better fit for the data than the line $y=\frac{5}{2}n+1$. [3]

Solutions

Consider $\mathrm{f}(n)=\frac{5}{2}n+c$.

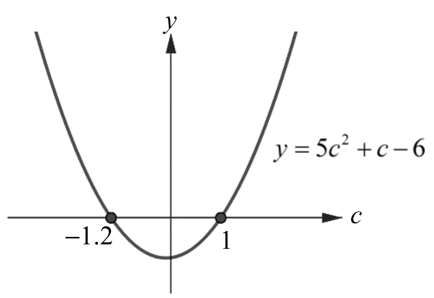
Point (a, b)	(2, 6)	(5, 10)	(6, 16)	(7, 20)	(11, 25)
$\mathrm{f}(a)$	$5+c$	$12.5+c$	$15+c$	$17.5+c$	$27.5+c$
Residual $b-\mathrm{f}(a)$	$1-c$	$-2.5-c$	$1-c$	$2.5-c$	$-2.5-c$

Sum of squares of residuals for the line $y=\frac{5}{2}n+c$ is

$(1-c)^2+(-2.5-c)^2+(1-c)^2+(2.5-c)^2+(-2.5-c)^2$

$=2(1-c)^2+2(2.5+c)^2+(2.5-c)^2$

$=20.75+c+5c^2$



For this to be a better fit,

$20.75+c+5c^2<26.75$

$5c^2+c-6<0$

Using GC, range of values of c is $\underline{-1.2<c<1}$

End of Paper