

[MAA 4.4] LINEAR REGRESSION

SOLUTIONS

Compiled by: Christos Nikolaidis

O. Practice questions

1. (a) II
 (b) V
 (c) III
 (d) I

2. (a)

	mean	standard deviation
for the values of x	5	2.55
for the values of y	2.5	1.12

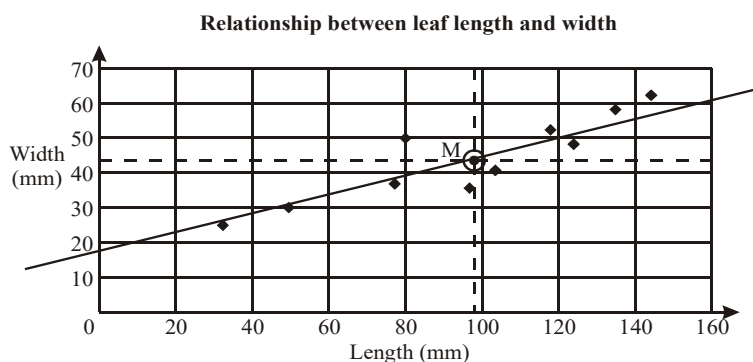
- (b)

L_1 : regression line of y on x	$y = 2.2x - 0.5$
L_2 : regression line of x on y	$x = 0.423y + 0.385$

- (c) (i) $r = 0.965$ (ii) strong positive
 (d) $y = 2.2x - 0.5 \Rightarrow y + 0.5 = 2.2x \Rightarrow x = 0.455y + 0.227$
 The result is different that L_2
 (e) (i) $y = 2.2 \times 5 - 0.5 = 10.5$
 (ii) If we use L_1 : $4 = 2.2y - 0.5 \Rightarrow y = 2.045$ (in fact we used the inverse function in (d))
 If we use L_2 : $x = 0.423 \times 4 + 0.385 \cong 2.077$
 The 2nd estimation (using the regression of x on y), is more reliable.

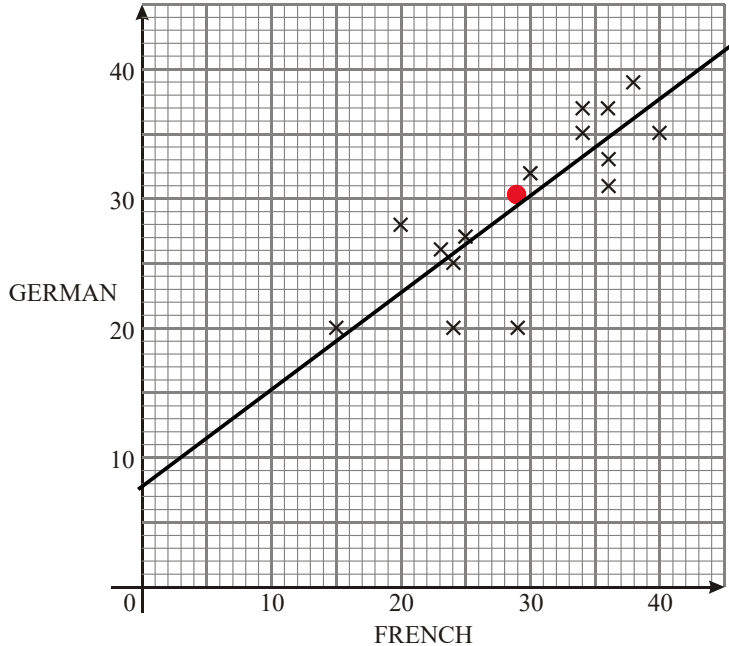
A. Exam style questions (SHORT)

3. (a) (i) ① (ii) ③
 (b) (i) ① 0.04 ② -0.20 ③ -0.85
 (ii) 1.60 A product-moment correlation coefficient cannot be > 1 .
 0.90 There is no diagram with a strong positive correlation.
4. (a) (see diagram)
 (b)

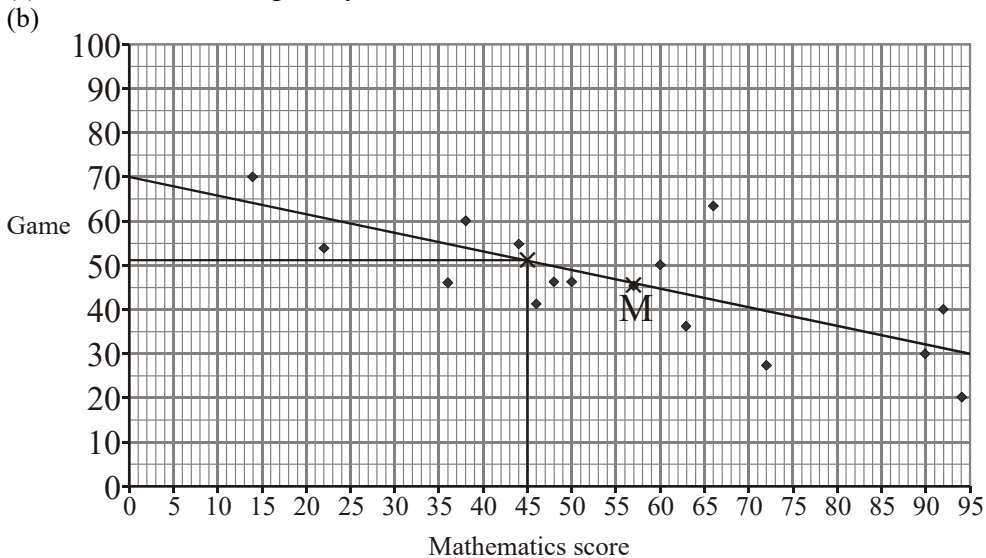


- (c) leaf length and leaf width are positively correlated

5. (a) $r = 0.76$
 (b) Fairly strong positive correlation between high school grades and university grades
 (c) $y = 0.052x - 1.29$ (3 s.f.)
6. (a) High positive *or* high *or* positive *or* good *correlation* etc.
 (b) M(29, 30)



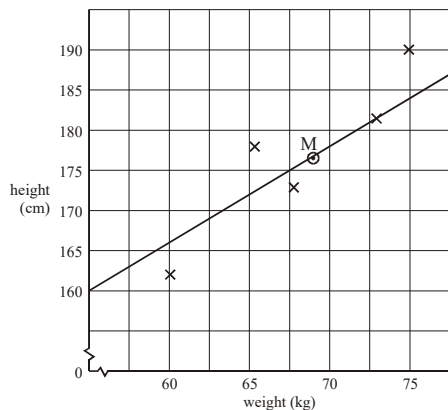
- (c) Suitable line which should pass through M and have nearly as many crosses (plotted points) below it as above it.
 (d) value obtained using candidate's line of best fit. (Follow through from part (c))
 For example, the line above gives $y = 32$ (or 33)
7. (a) The scores are negatively correlated



- Line must be drawn straight.
 It must pass through (0, 69).
 It must pass through the mean point M = (56.9, 45.9).
 (c) 51 is closest. Allow 50 or 52 (ft from candidate's graph).

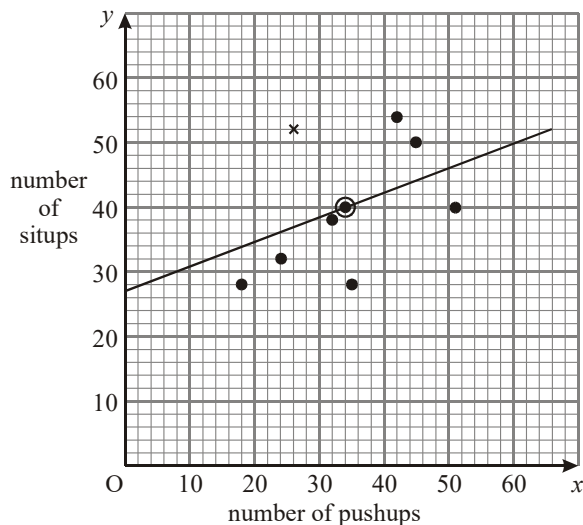
8. (a) (i) $\frac{182+173+162+178+190}{5} = 177\text{cm}$
(ii) $\frac{73+68+60+66+75}{5} = 68.4\text{ kg}$
(Or directly by GDC)

(b)



(c) $r = 0.943$

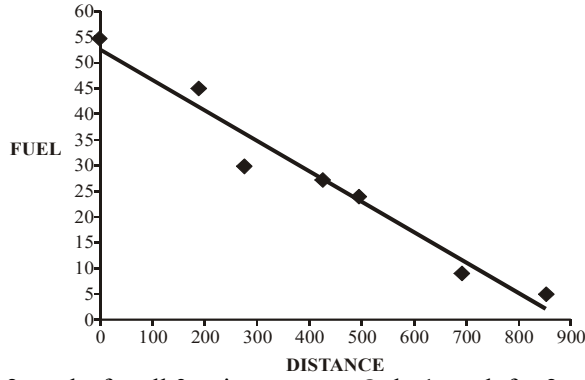
9. (a) On the graph
(b) $y = 0.373x + 27.3$
(c) $\bar{x} = 34$ and $\bar{y} = 40$



- (c) $y = 0.373 \times 60 + 27.3 \approx 49.68$
About 50 situps
(or by using the graph of the candidate, allow ± 2)

10. (a) $r = 0.6399706... \approx 0.64$ (2 d.p.)
(b) This indicates that there is a degree of positive correlation between scores in Mathematics and scores in English.
Therefore those who do well in Mathematics are likely to do well in English also. (Or equivalent statements.)

11. (a)



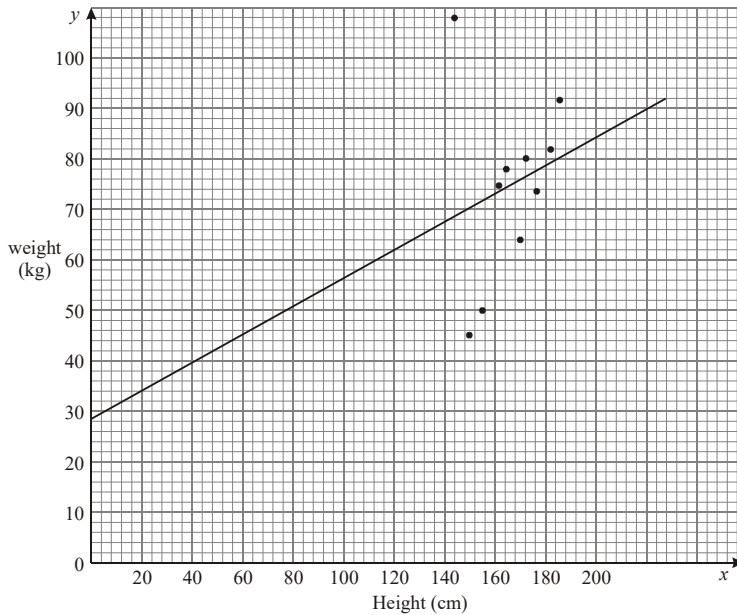
2 marks for all 3 points correct. Only 1 mark for 2 points correct

- (b) Straight line **with negative gradient** passing through the mean intercept on y-axis between 50 and 55
- (c) 32 (read answer from candidate's line)

B. Exam style questions (LONG)

12. (a)

x	155	161	173	150	182	165	170	185	175	145
y	50	75	80	46	81	79	64	92	74	108



- (b) Mean height = $166.1 = 166$ (3 s.f.)
Mean weight = 74.9 (3 s.f.)
- (c) (i) $y = 0.276x + 29.1$
(ii) Line on graph. y-intercept at 29.1, straight line through (166, 74.9).
- (d) (i) $y = 0.276 \times 190 + 29.1 = 81.5$ kg
(ii) $72 = 0.276x + 29.1$
$$x = \frac{72 - 29.1}{0.276} = 155 \text{ cm.}$$

OR From the graph
- (i) $y = 81 (\pm 1)$ (ii) $x = 155 (\pm 1)$

- (e) The 'line of best fit' becomes closer to the remaining points.
OR
 Gradient becomes steeper and the line is more accurate 'best fit'.
OR
 Any reasonable explanation. (Line becomes $y = 1.10x - 113$)

13. (a) $y = 0.070x - 3.22$

Accept $0.07x$.

- (b) for each cm of height the shoe size increases by 0.070

(c) (i) $y = 0.070 \times 162 - 3.22 = 8.12$

Therefore shoe size 8 or 9 (8.12).

OR $y = 8$ or 9

(ii) interpolation since a.62 is within the range of values of x .

(d) $r = 0.681$

- (e) Moderately strong, positive correlation.

14. (a) (i) $r = 0.935$ (3 s.f.)

(ii) it suggests a strong positive correlation between the two variables.

(b) $y = 0.291x + 1.56$

(c) (i) $y = 0.291 \times 30 + 1.56 = 10.29 = 10$ hours

(ii) $8 = 0.291x + 1.56$

$x = 22.13 = 22$ years

(d) $x = 3y - 0.846$, For $y = 8$, $x = 23.154 = 23$

15. (a) (i) 1992 mean = \$1.59, Sd = \$0.727 (or 0.73)

(ii) 2002 mean = \$1.98, Sd = \$0.635 (or 0.64)

(b) (i) $r = 0.672$

(ii) There is a **weak positive** correlation

(c) $y = 0.588x + 1.05$

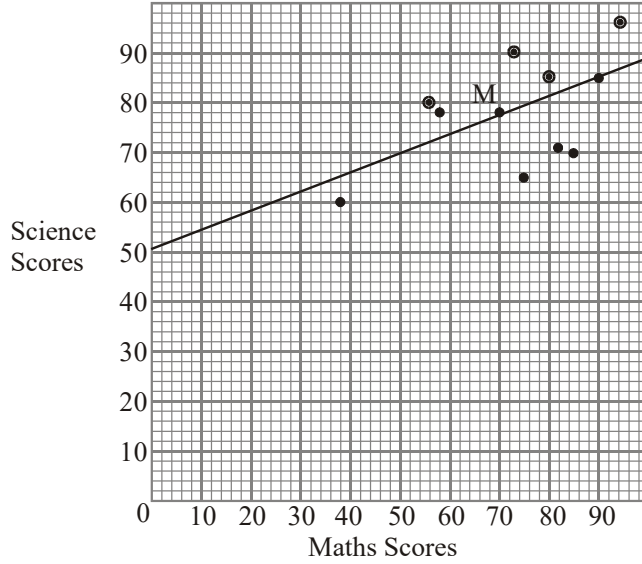
(d) $y = 0.588 \times 2.60 + 1.05 = \2.58

- (e) Coffee – because it is the only item to go down in price.

OR

Rolls – because the price increased significantly.

16. (a)



(b) Point M(73,78)

(c) $y = 0.359x + 51.8$

(d) going through M, y intercept anywhere from 50 to 54

(e) $y = 0.359 \times 88 + 51.8 = 83$

OR

$y = 83 (\pm 2)$ if read from the graph and method is shown.

17. (a) $y = 0.290x + 56.5$

(b) $x = 0.770y + 14.3$

(c) $r = 0.472$

(d) $r_s = 0.444$