EXERCISES [MAI 4.4]

LINEAR REGRESSION

SOLUTIONS

Compiled by: Christos Nikolaidis

A. Paper 1 questions (SHORT)

1. (a) 0.965

(b) strong positive

(c) $y = 2.2x - 0.5$

(d) whenever x increases by 1 unit, y increases by 2.2 units.

(e) The value of y corresponding to 0 units of x.

2. (a) $y = 2.2x - 0.5 \Leftrightarrow y + 0.5 = 2.2x \Leftrightarrow x = 0.455\, y + 0.227$

(b) $x = 0.423y + 0.385$

(c) The relation in (a) is in fact the inverse function of the line $y = 2.2x - 0.5$
If y is given, the answer in (c) gives a more reliable estimation of x.

3. (a) $r_s = 1$.

(b) It describes the monotonic relationship of the data. When x increases y also increases.

4. (a) II

(b) V

(c) III

(d) I

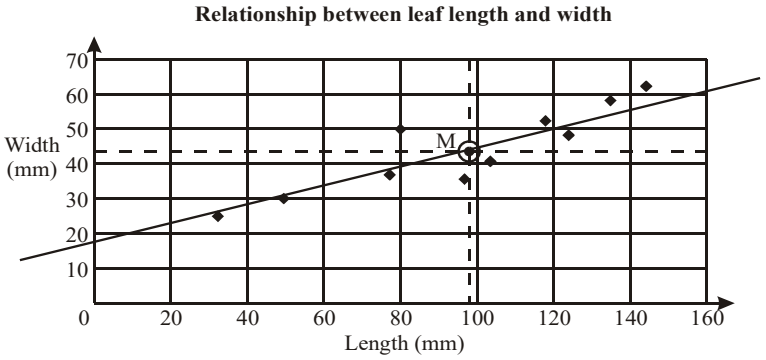
5. (a) (i) ① (ii) ③

(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.

6. (a) (see diagram)

(b)



(c) leaf length and leaf width are positively correlated

7. (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.)

8. (a) High positive ***or*** high ***or*** positive ***or*** good *correlation* etc.

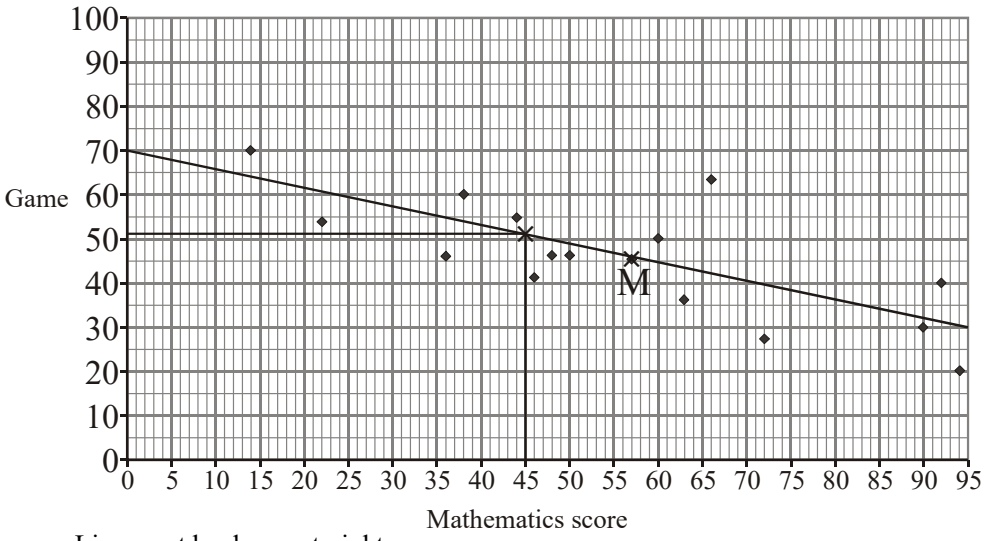
(b) M(29, 31)

(c) Suitable line which should pass through the candidate's M and have nearly as many crosses (plotted points) below it as above it.

(d) value (including non-itegers) obtained using candidate's line of best fit. (Follow through from part (c).)

9. (a) The scores are negatively correlated

(b)



Line must be drawn straight.

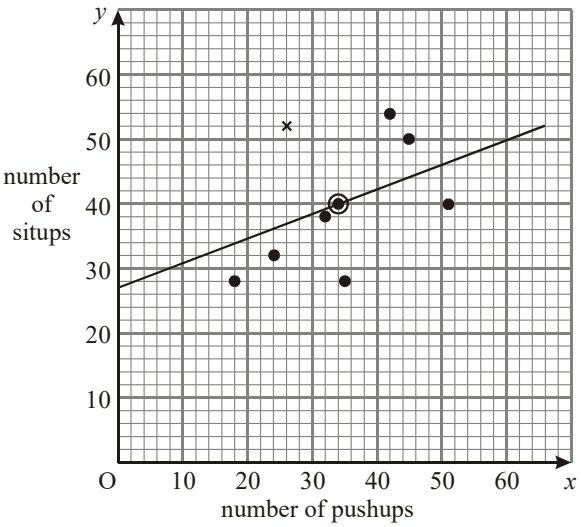
It must pass through (0, 69).

It must pass through the mean point M = (56.9, 45.9).

(c) 51 is closest. A1low 50 or 52 (**ft** from candidate's graph).

10. (a) On the graph

(b) $\bar{x} = 34$ and $\bar{y} = 40$



(c) 50 situps (allow ±2) (**ft** from candidate's graph)

11. (a) $r = 0.6399706\ldots \approx 0.64$ (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.)

12. (a) $r = 0.838$

(b) it is the average of the ranks 4,5,6,7

(c)

(x)	(y)	Rank x		Rank y	
		original	modified	original	modified
10	100	1	1	1	1
20	110	2	2	2	2
30	130	3	3	4	4.5
40	120	4	5.5	3	3
40	130	5	5.5	5	4.5
40	140	6	5.5	6	6.5
40	150	7	5.5	8	8
50	140	8	8	7	6.5

Spearman rank correlation coefficient $r_s = 0.761$

B. Paper 2 questions (LONG)

13. (a) $r = 0.929$

(b) strong positive

(c) $y = 0.929\,x + 10.3$

(d) $x = 0.929\,y - 9$

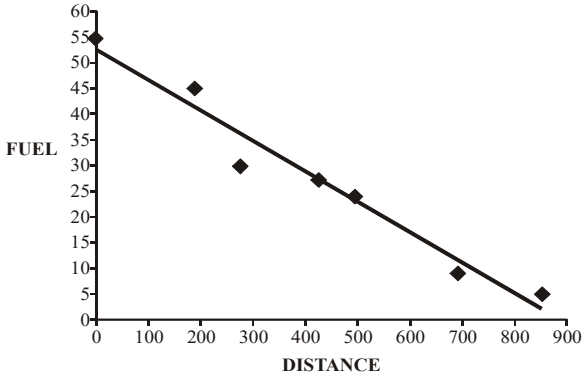
(e) ranks

x	y
1	1
2	3
3	2
4	4
5	6
6	5

$r_s = 0.886$

(f) $r = 0.929$ indicates the degree of linear relationship between x and y (strong positive)
$r_s = 0.886$ indicates the degree of monotonic relationship between x and y (in what extent y increases when x increases)

14. (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)

(d) $r = -0.978$

(e) ranks

x	1	2	3	4	5	6
y	6	5	4	3	2	1

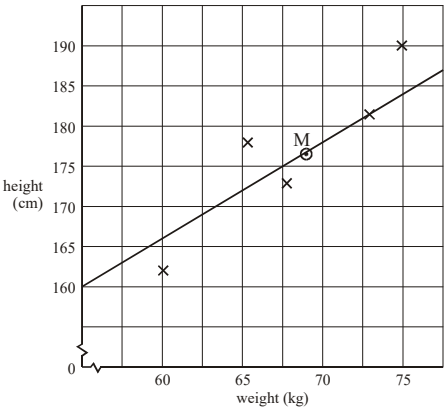
$r_s = -1$

(f) $r = -0.978$ indicates the degree of linear relationship between x and y (strong negative)
$r_s = -1$ indicates a perfect negative monotonic relationship between x and y
(y decreases when x increases)

15. (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) $r = 0.943$

(d) ranks

162	60	1	1
173	68	2	3
178	66	3	2
182	73	4	4
190	75	5	5

$r_s = 0.9$

(e) Because the rank of y for 173 is larger than the rank of y for 178 (no perfect monotonic relationship)

16. (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

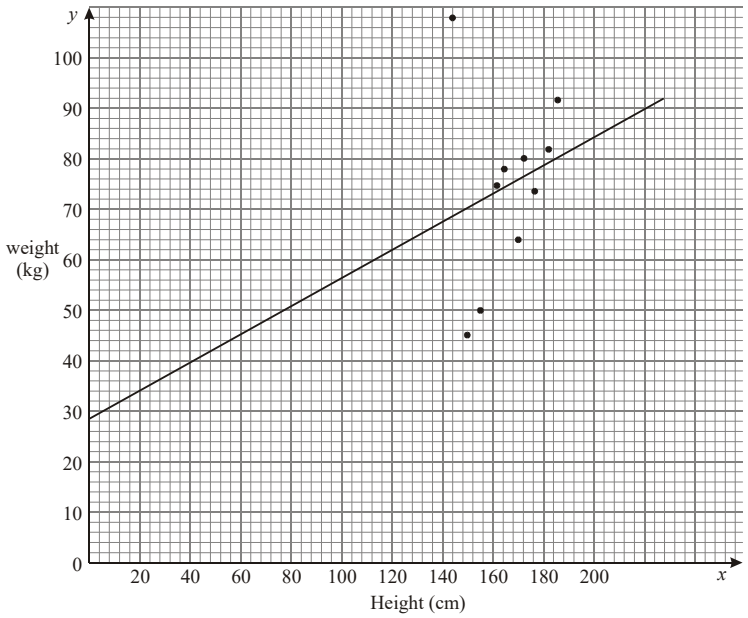
(ii) interpolation since a.62 is withing the range of values of x.

(d) $r = 0.681$

(e) Moderately strong, positive correlation.

17. (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) $y = 81$ (±1)

(ii) $x = 155$ (±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 $y = 1.10x - 113$)

18. (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 $y = 8$, $x = 23.154 = 23$

19. (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

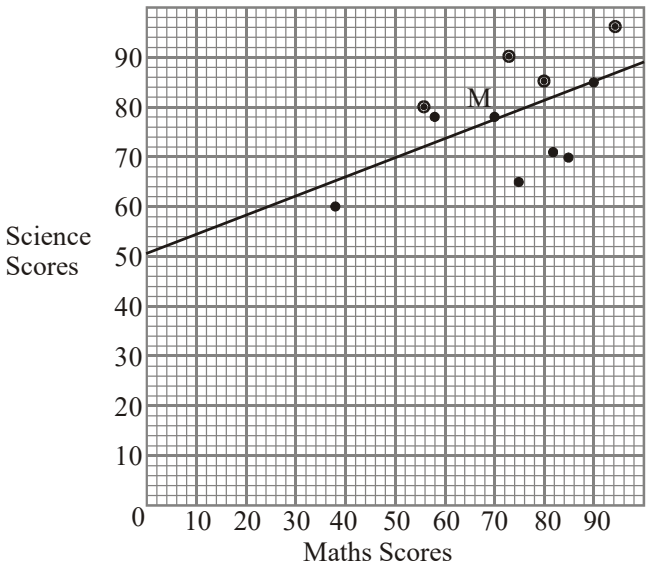
(d) $y = 0.588 \times 2.60 + 1.05 = \$2.58$

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

OR

Rolls – because the price increased significantly.

20. (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

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

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

(b) $y = 0.770x + 14.3$

(c) $r = 0.472$

(d)

Mathematics (x)	**Science (y)**	Rank x	Rank y
38	60	1	1
56	80	2	6
58	78	3	5
73	65	**4 →4.5**	2
73	90	**5 →4.5**	9
80	85	6	7 →7.5
90	70	**7 →8**	3
90	71	**8 →8**	4
90	85	**9 →8**	8 →7.5
95	96	10	10

$r_s = 0.444$