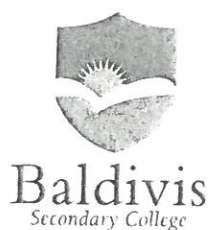


## BALDIVIS SECONDARY COLLEGE



### APPLICATIONS - Unit 3 & 4

#### Test 1- Bivariate Data

**SOLUTIONS**

Student Name \_\_\_\_\_ Teacher Name \_\_\_\_\_

**Time allowed for this task:** 55 minutes, in-class, test conditions.

Section 1: 20 minutes

Section 2: 35 minutes

**Materials required:**

Section 1 Calculator free section (15 marks)  
Standard writing equipment  
SCSA Formula Sheet

Section2 Calculator assumed section (35 marks)  
Calculator (to be supplied by the student)  
SCSA formula Sheet  
One page (doubled sided) of hand written notes

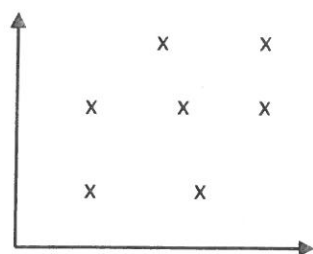
**Other materials allowed:** Drawing templates

**Marks available:** 50 marks

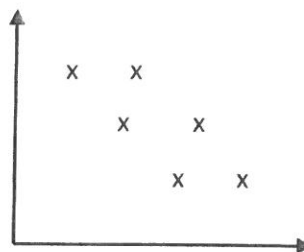
**Task Weighting:** 5%

**Question 1. (3 marks)**

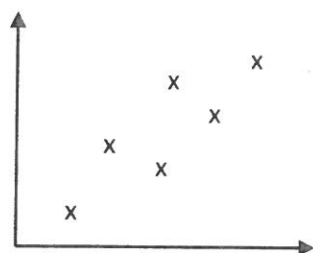
Consider the scatter graphs shown.



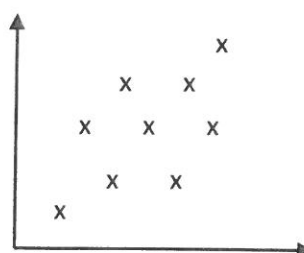
Graph A



Graph B



Graph C



Graph D

(a) Match each graph with an appropriate correlation coefficient given below.

(2)

Correlation coefficient	0.2	0.5	-0.8	0.8
Graph	A	D	B	C

(b) If each scattergraph above had a line of prediction fitted, and an interpolation was made, which of the four predicted values would be the most unreliable?

(1)

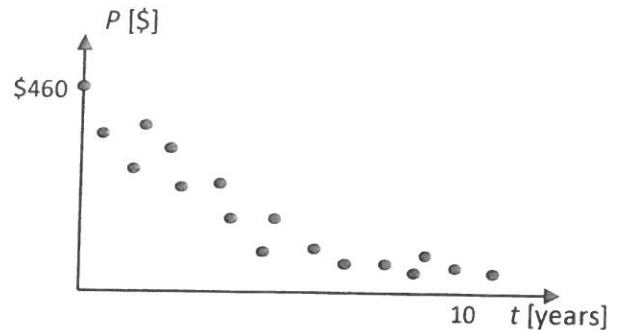
Graph A

**Question 2. (5 marks)**

The graph to the right shows the market value of a popular product over a period of 10 years.

The correlation coefficient for this graph is:

$$r_{tp} = -0.8793$$



- (a) Describe the relationship of the data based on the correlation coefficient.

(1)

As the years progress, the value of the product decreases

- (b) Which of the least squares regression lines given below would be the most suitable for this graph? Circle the correct answer.

(1)

(i)  $P = 32.4t + 460$

(ii)  $P = -32.4t - 460$

(iii)  $P = -32.4t + 460$  \*

- (c) It is suggested that if we attempt to predict the market value of this product well beyond the 10 year period, e.g. 15+ years, the linear regression model could be unsuitable. Comment on the validity and reasoning behind this statement.

(2)

Invalid because using extrapolation, with the regression line equation, the value of the product would be negative which is not possible.

- (d) We are required to predict reliable market values of this product after the 10 year period. What model should be used? Explain.

(1)

A parabolic/exponential/curve model as the rate of decrease of the product value (as time increases) is slowing.

**Question 3. (4 marks)**

Consider the following statement.

*"The coefficient of linear correlation between Biology marks and History marks is 0.94"*

(a) Does this statement imply that students who are good in Biology are also good in History?

Justify your answer.

<sup>Yes</sup>  
~~No~~ It means that if a student is picked at random it most likely that their marks will be similar but we cannot be certain that this will always be the case.

(b) Does this statement imply that by improving my Biology marks I will notice an improvement in my History marks?

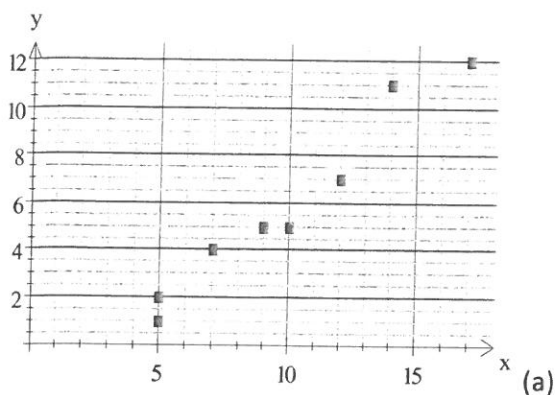
Justify your answer.

No

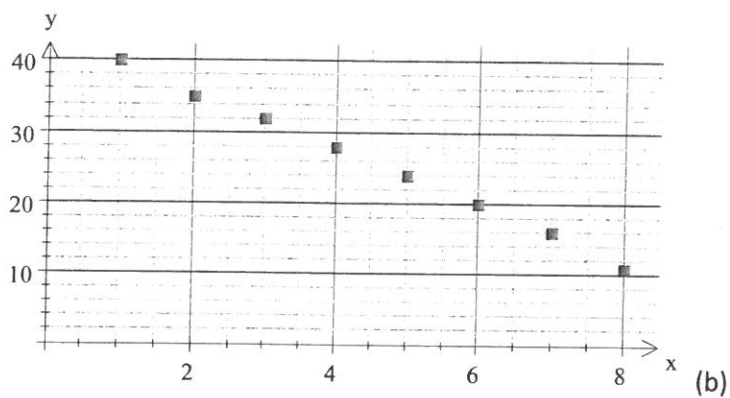
There is a correlation but not necessarily causation, both may be caused by another factor

**Question 4. (3 marks)**

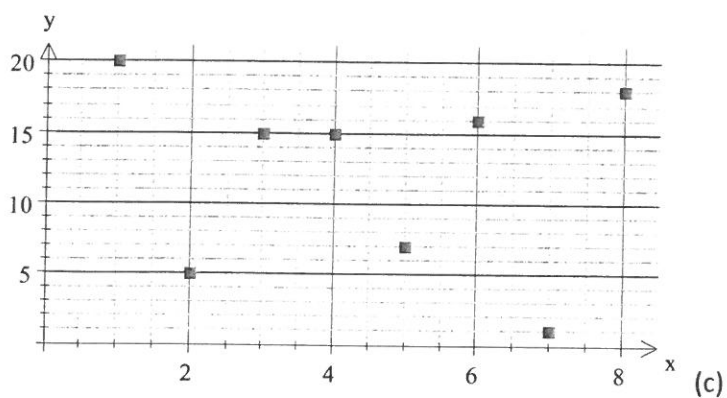
Describe the correlation of the following scattergraphs.



Strong Positive



Perfect Negative



No correlation

## Section 2 - Calculator Allowed

Total marks – 35

Working time 34 minutes

### Question 5. (7 marks)

The following data was collected by a student interested in buying a second-hand sailing boat. It shows the age in years and the sale price, in hundreds of dollars, for 10 boats of the same type.

Age ( $t$ )	3	13	5	14	5	10	6	12	4	9
Price ( $p$ )	41	24	34	18	29	31	32	23	35	28

- a) Use your calculator to graph this data and by referring to features of your graph and a suitable statistic, explain why it is appropriate to fit a linear relationship to model the price of these boats against time. [2]

A linear relationship is appropriate as the data generally forms a straight line.

It is appropriate to model price against time as we are interested to know the change in price over the years

- b) Calculate the least-squares linear regression line of  $p$  on  $t$ . [2]

$$p = -1.5t + 41.66$$

- c) The student saw another boat aged 15 years old advertised in the local paper. Use your line from b to predict its sale price and comment on the reliability of your prediction. [3]

$$t = 15$$

$$\therefore p = -1.5(15) + 41.66$$

$$p = -22.5 + 41.66$$

For a 15yr old boat  
predicted price is \$33.82

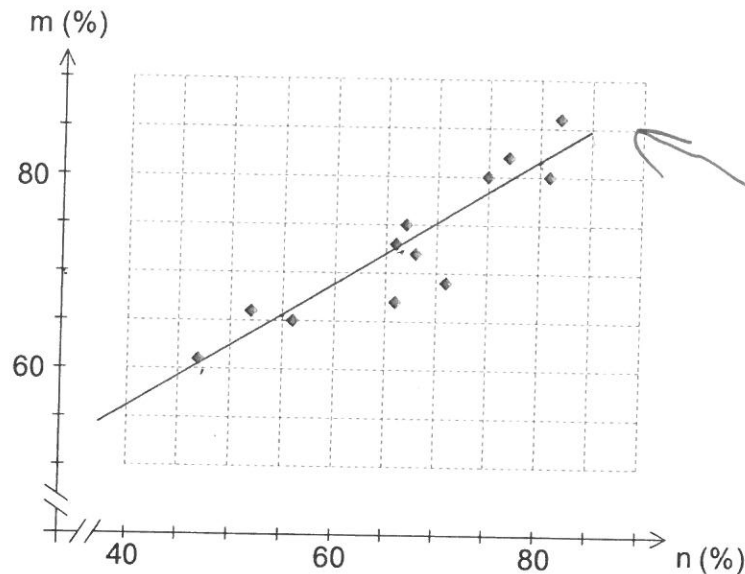
The price predicted is only somewhat reliable.

While the correlation is fairly strong, the prediction uses extrapolation which decreases this reliability.

### Question 6 ( 10 marks)

The table and scatterplot show the percentage scores for a group of twelve candidates who took both numerical reasoning (n) and mechanical aptitude (m) tests as part of a job selection process.

n (%)	m (%)
47	61
82	86
67	75
66	67
71	69
56	65
68	72
66	73
52	66
81	80
77	82
75	80



- (a) Determine the least-squares regression line of  $m$  on  $n$ .

$$m = 0.64n + 30$$

- (b) Draw your line from (a) on the scatterplot above.

- (c) State the correlation coefficient,  $r_{nm}$ , for the linear model in (a).

0.91

- (d) Describe two features of the scatterplot above that can be determined from  $r_{nm}$ .

- Points must be bunched together
- Unlikely to be an outlier

(other reasonable answers accepted)

- (e) A thirteenth candidate for the job scored 45 on the numerical reasoning test but due to illness was unable to complete the mechanical aptitude test. What mechanical aptitude score would you predict for this person? Explain, with reasons, how valid your prediction is.

(3)

$$n = 45$$

$$\therefore m = 0.64(45) + 30$$

$$m = 59$$

Prediction for  $m\%$  score = 59%

The price is somewhat reliable.

Strong positive relationship but  
extrapolation.

### Question 7 (15 marks)

Suppose a survey was carried out to investigate whether there was any evidence to suggest that the likelihood of someone developing heart problems in their teenage years was associated with the amount that their mother engaged in a particular activity during pregnancy. Further suppose that the survey produced the following results.

		Level of heart problems in teenage years		
		None	Moderate	Severe
Engaged in activity during pregnancy	Not at all	70	18	6
	Sometimes	20	14	5
	Often	12	15	3

- a) What is the explanatory variable and which is the response variable in this situation? [2]

Exp: Engaged in activity while pregnant

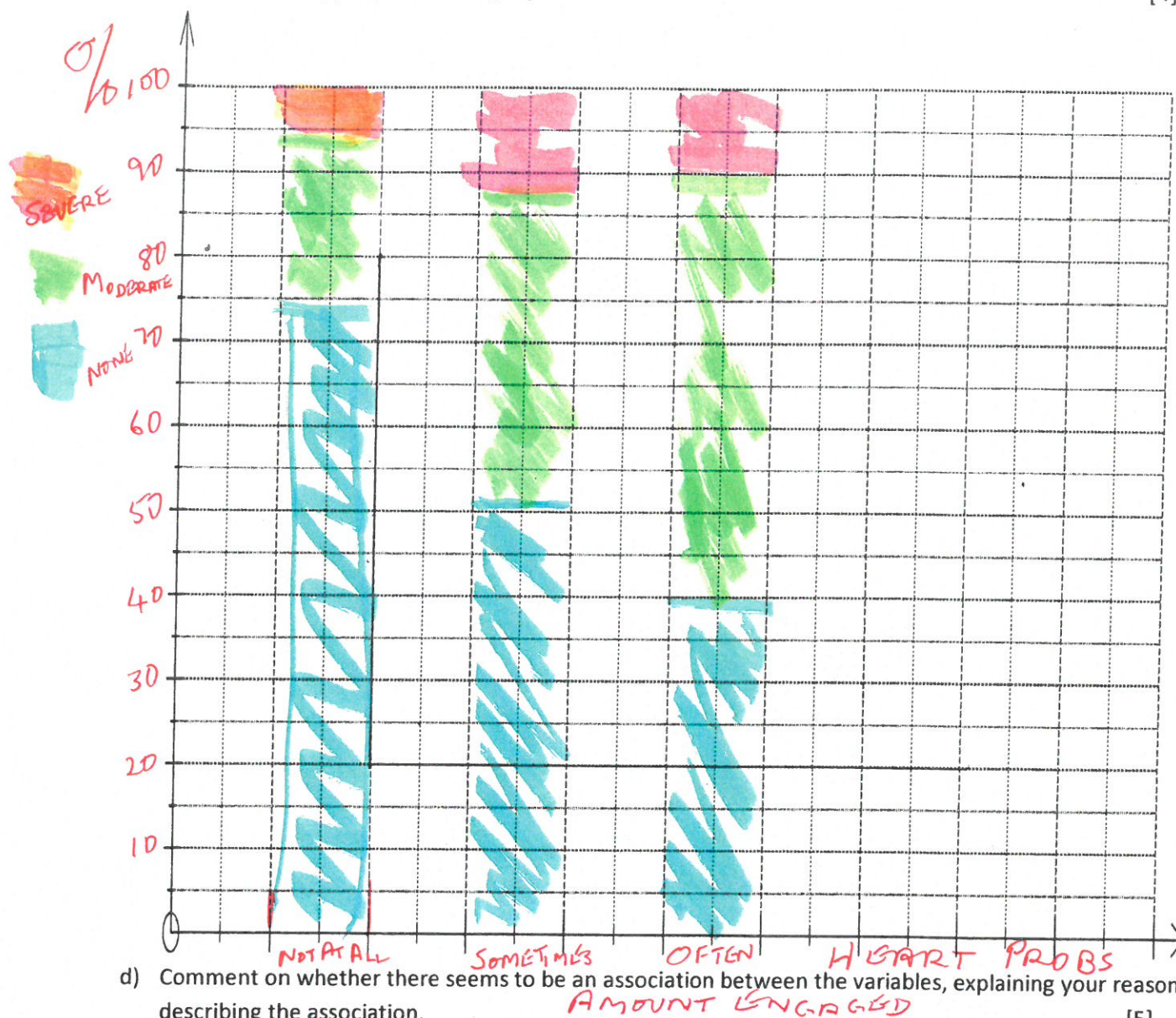
Resp: Level of heart problems in teenage yrs

- b) Recreate the table showing either row or column percentages as appropriate. [4]

		Heart Problems in Teenage yrs		
		None	Moderate	Severe
Engaged in Activity During Pregnancy	Not at all	74%	19%	6%
	Some times	51%	36%	13%
	Often	40%	50%	10%



- c) Use your table to create a proportional column graph with equal height columns each representing 100% of an appropriate category. [4]



- d) Comment on whether there seems to be an association between the variables, explaining your reasons and describing the association. [5]

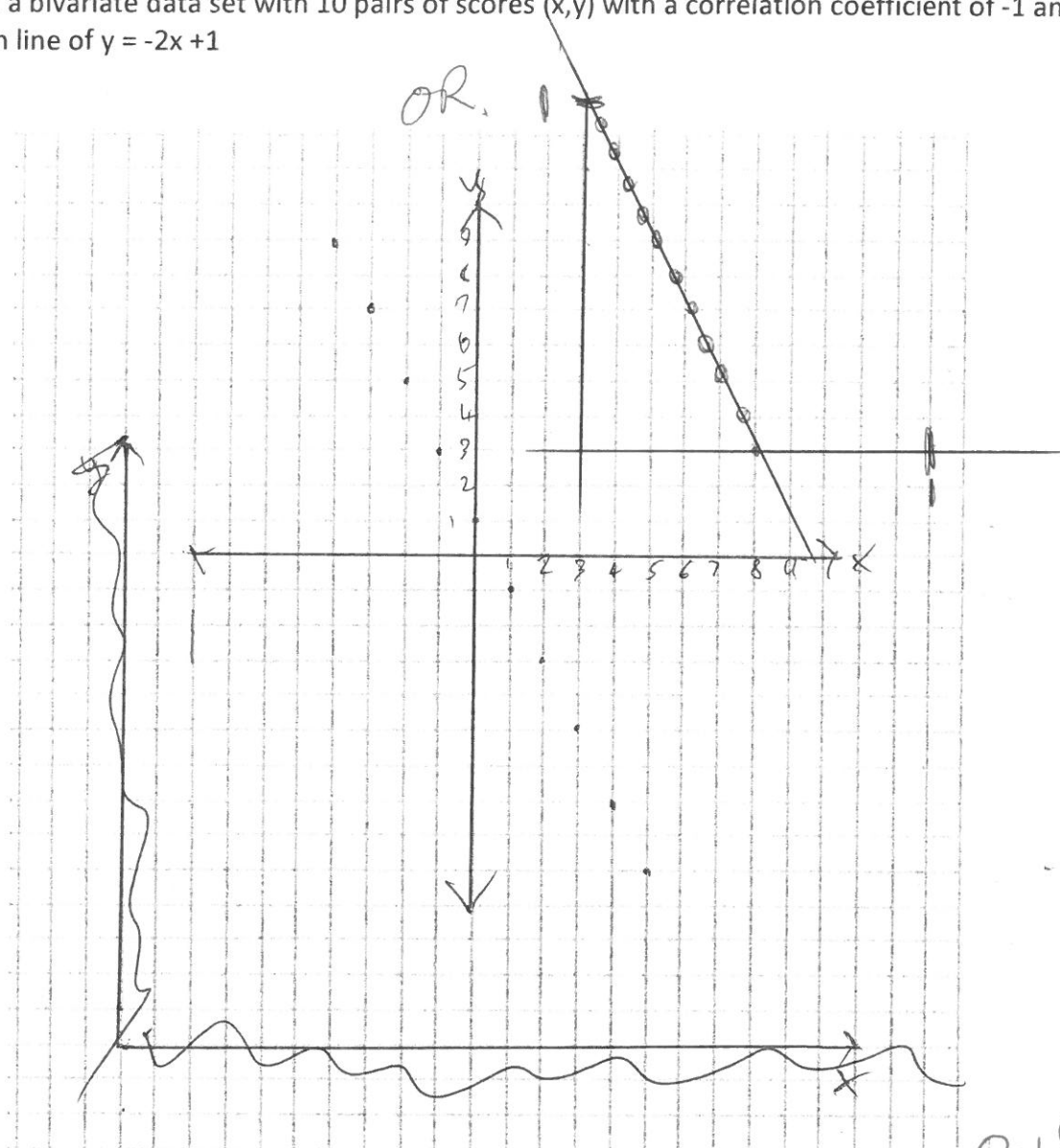
*SOMETHING LIKE*

Yes, there seems to be a correlation, as the amount that the pregnant mother engaged in the activity increases from not at all to often, the risks of a teenager suffering no heart problems decreases from 74.5% to 40%, and the likelihood of a teenager suffering moderate heart problems increases from 19% to 50%. Only the severe category does not fully correlate. Where a woman who did not engage in the activity at all gives her child a 6.4% chance, to a woman who sometimes engaged doubles to a 12% chance, while the woman who engaged often often is a 10% chance.

End of test

**Question 8 (3 marks)**

Construct a bivariate data set with 10 pairs of scores  $(x, y)$  with a correlation coefficient of -1 and a regression line of  $y = -2x + 1$



SHOW SCALES ON GRAPH ✓  
DOTS ON ST LINE ✓  
LINE GRAPH ✓