**BALDIVIS SECONDARY COLLEGE**

** APPLICATIONS - Unit 3 & 4**

**Test 1- Bivariate Data**

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.

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

x

Graph A

Graph B

Graph C

Graph D

1. Match each graph with an appropriate correlation coefficient given below. (2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Correlation coefficient | 0.2 | 0.5 | –0.8 | 0.8 |
| Graph |  |  |  |  |

1. 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)

**Question 2. (5 marks)**

10 *t* [years]

*P* [$]

$460

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:

*rtP* = –0.8793

1. Describe the relationship of the data based on the correlation coefficient. (1)
2. Which of the least squares regression lines given below would be the most suitable for this graph? Circle the correct answer. (1)

(i) *P* = 32.4*t* + 460

(ii) *P* = –32.4*t* – 460

(iii) *P* = –32.4*t* + 460

1. 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)
2. We are required to predict reliable market values of this product after the 10 year period. What model should be used? Explain. (1)

**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.

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

Justify your answer.

**Question 4. (3 marks)**

Describe the correlation of the following scattergraphs.

(a)

(b)

(c)

End of non-calculator Section.

 **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 |

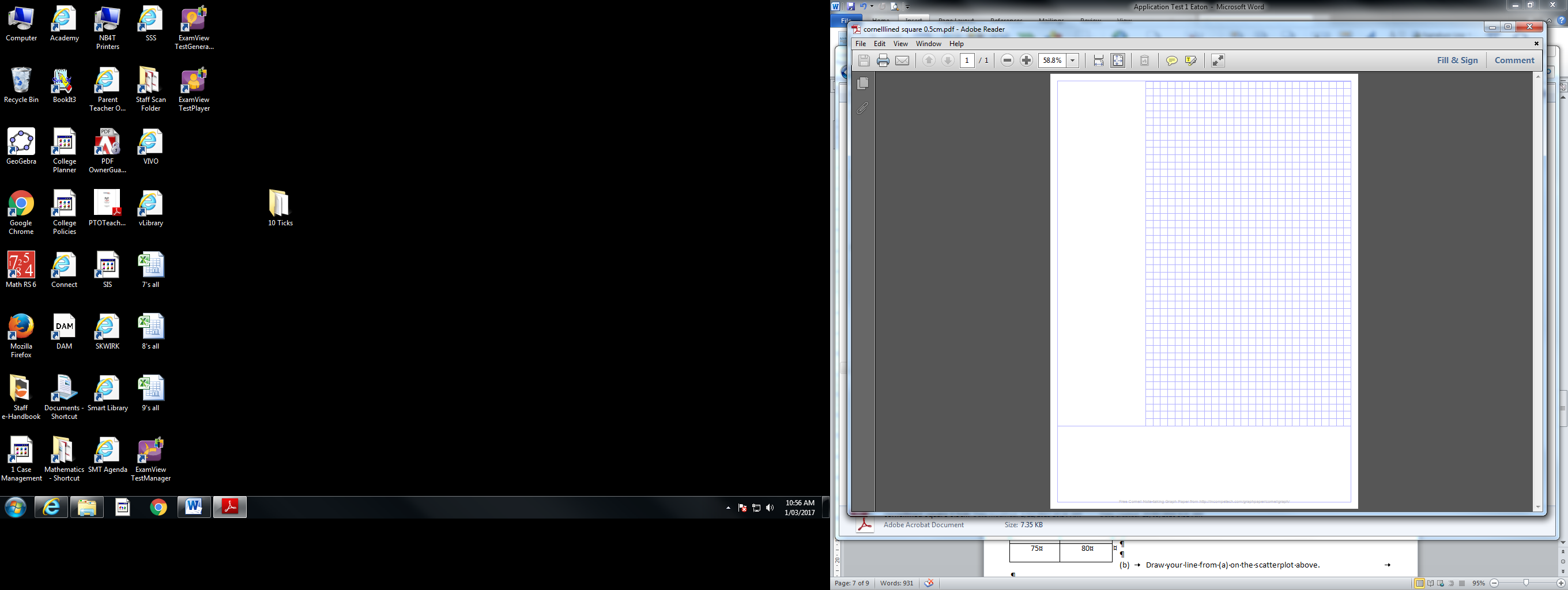
a) Use your calculator to graph this data and by referring to features of your graph and a suitable statistic, explain why it is appropriate to fit a linear relationship to model the price of these boats against time. [2]

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

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]

**Question 5 (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



**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*. (2)

(b) Draw your line from (a) on the scatterplot above. (2)

(c) State the correlation coefficient, *rnm*, for the linear model in (a). (1)

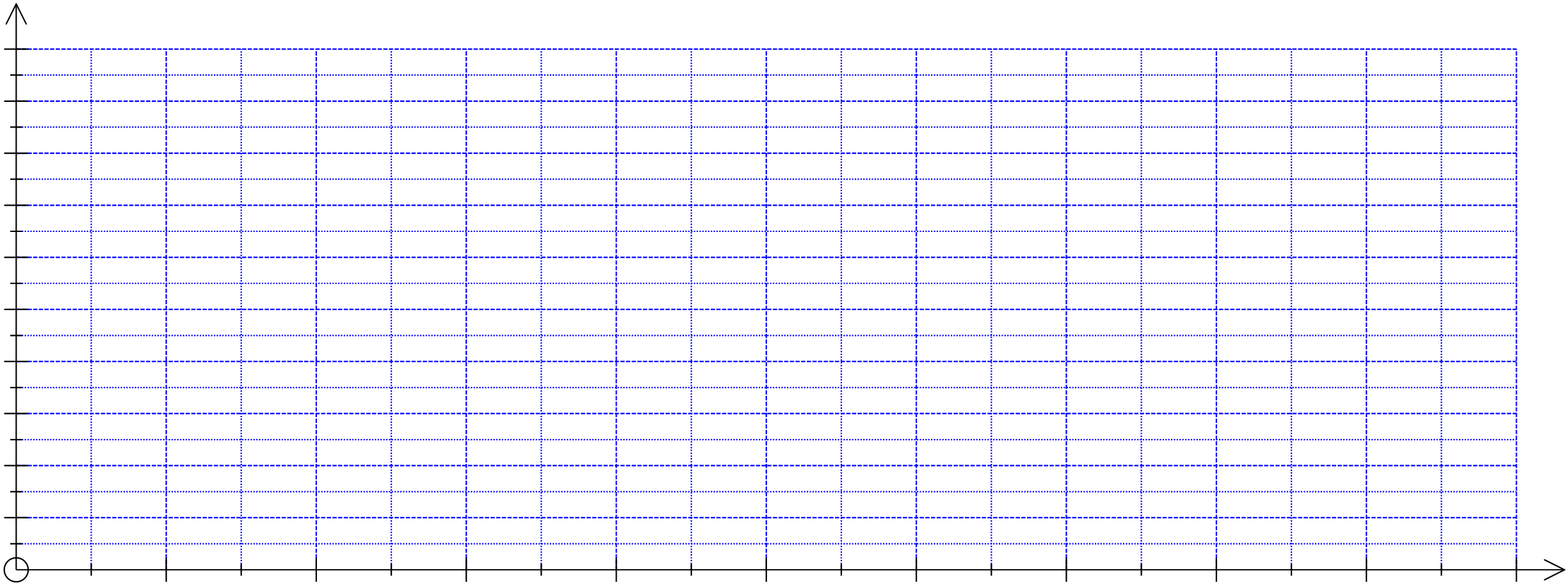
(d) Describe two features of the scatterplot above that can be determined from *rnm*, (2)

(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)

**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 |

1. What is the explanatory variable and which is the response variable in this situation? [2]
2. Recreate the table showing either row or column percentages as appropriate. [4]
3. User your table to create a proportional column graph with equal height columns each representing 100% of an appropriate category. [4]
4. Comment on whether there seems to be an association between the variables, explaining your reasons and describing the association. [5]

End of test