



Greenwood College
Year 12 Applications
Test 2 2019
Resource-Free

Name.....

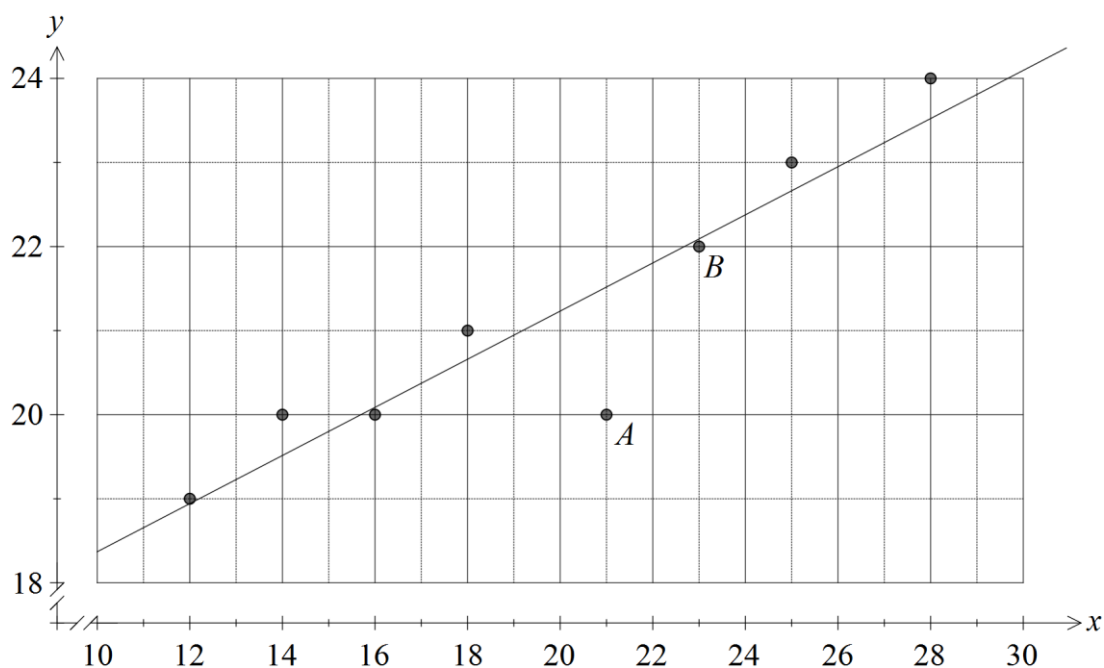
No calculators nor notes allowed.
 28 mark total.

Formula sheet allowed.
 30 minute time limit.

Question 1

[8 marks: 2, 1, 1, 3, 1]

The scatterplot below, with least-squares line displayed, shows the relationship between two numerical variables, x and y . The correlation coefficient between the variables is 0.92.

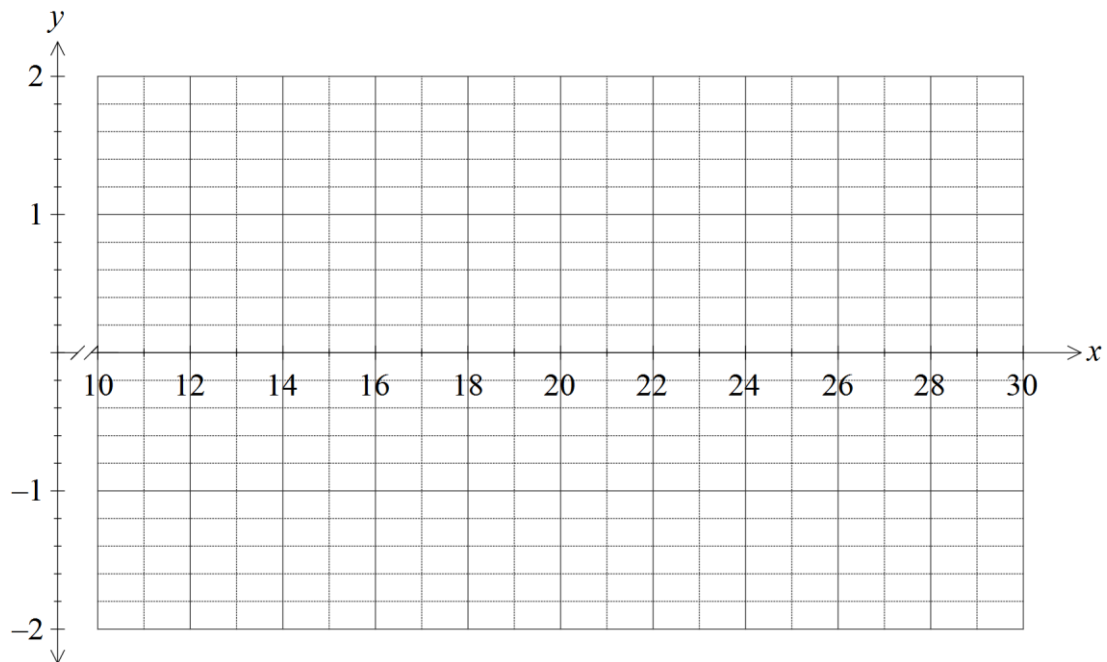


- (a) Describe the association between x and y in terms of direction and strength.

- (b) Describe the effect on the correlation coefficient if
- (i) the point labelled *A* was removed from the dataset.

- (ii) the point labelled *B* was removed from the dataset.

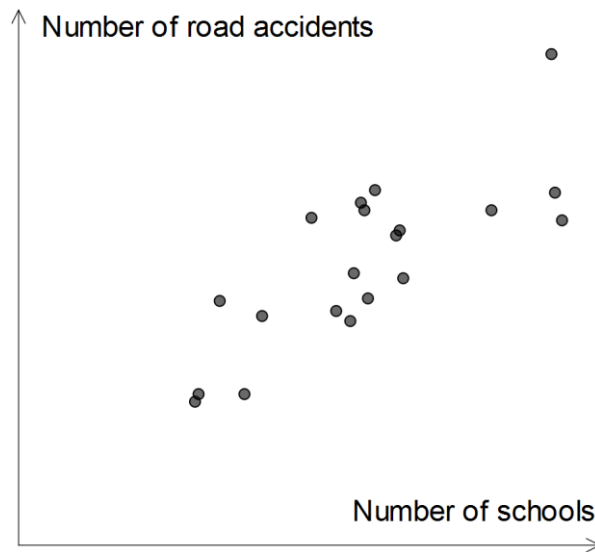
- (c) Sketch a residual plot for the eight paired values on the axes below.



- (d) Comment, with reasons, on the appropriateness of fitting a linear model to this dataset.

Question 2**[5 marks: 2, 1, 2]**

The scatterplot below shows data from a sample of towns in a region.



- (a) Which of the numbers 1, -1, 0.75, -0.75, 0.5, -0.5, 0.25, -0.25 and 0 is closest to the correlation coefficient between the two variables? Explain your choice.
- (b) A politician saw the graph and claimed the data supported his plan to merge small schools and hence reduce the number of schools in individual towns. Identify a reason the politician might have had to make such a claim.

Question 2 cont.

- (c) Identify and explain a possible non-causal explanation for the observed association between the number of schools and the number of road accidents in this sample of towns.

Question 3**[8 marks: 2, 1, 1, 2, 2]**

- (a) Describe a suitable method to organise and display data when investigating the existence of an association between two categorical variables.

- (b) A class was set a task to investigate whether an association exists between the distance a student lived from school and the number of times they were late in a term.
 - (i) What **type** of graph/s would be appropriate to display data collected?

 - (ii) What statistical measure would be useful to calculate in order to determine whether an association existed?

Question 3 cont.

- (iii) One student designed the questionnaire shown below. Comment on the appropriateness of their design for this investigation.

Name:		
<i>Tick one box</i>	Distance less than 2 km	Distance more than 2 km
Late less than 3 times		
Late more than 3 times		

- (iv) A student carried out the investigation, found that a moderate negative association existed, and concluded that frequent lateness was caused by living close to the school. Comment on their conclusion.

Question 4**[5 marks: 1, 2, 2]**

A group of university students was asked the question “Does full attendance at school lead to an improved examination result?”

The results are summarised below.

	Agree	Disagree	Undecided
Under 20	14	42	14
20-25	56	16	8
Over 25	42	5	3

(a) State the explanatory variable for these data.

(b) The results are now presented as row percentages.

	Agree	Disagree	Undecided
Under 20	20%	60%	20%
20-25	70%	20%	10%
Over 25	84%	10%	6%

Show how the 20% (highlighted in **bold**) was calculated from the first table.

(c) Use the data to determine one association between the variables. Describe the association and explain your reasoning.

Question 5**[2 marks]**

Given a correlation co-efficient of 0.8, determine what percentage of the variation in response variable can be explained by the variation in the explanatory variable.



Greenwood College
Year 12 Applications
Test 2 2019
Resource-Allowed

Name.....

Formula sheet, one A4 page single-sided of notes and calculators allowed.

23 mark total.

25 minute time limit.

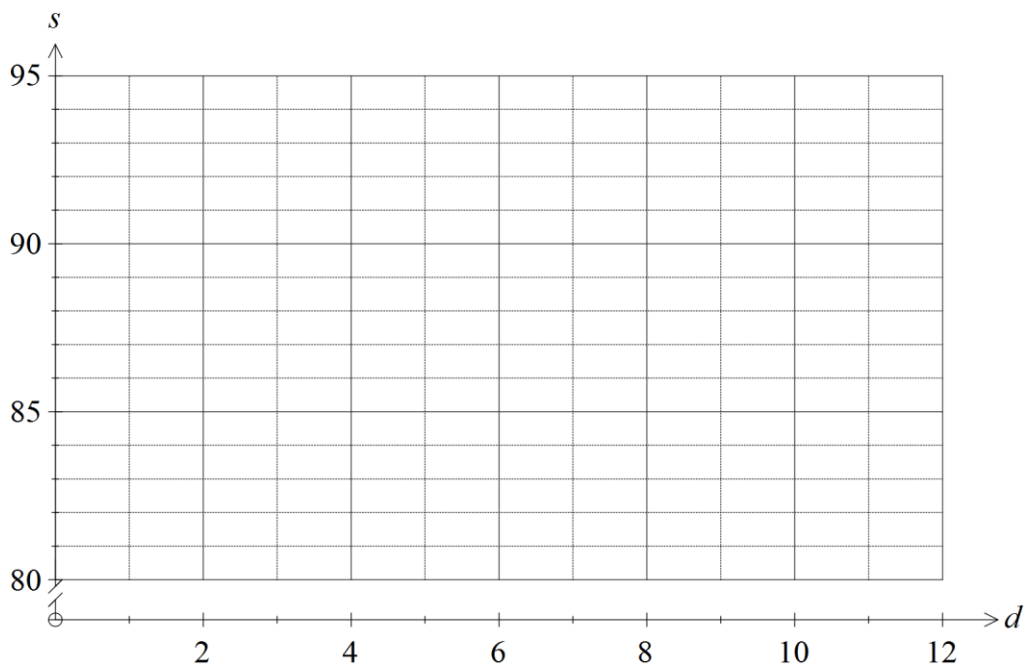
Question 6

[12 marks: 2, 3, 2, 1, 1, 1, 2]

The daily customer satisfaction index was measured by an online business over a period of ten consecutive days and the data collected is shown in the table below.

Day (d)	1	2	3	4	5	6	7	8	9	10
CS Index (s)	92.1	91.2	90.6	88.9	88.1	87.7	87.4	86.6	85.4	85.1

(a) Plot the above data on the axes below.



Question 6 cont.

- (b) Determine the equation of the least-squares line that models the linear relationship between the day number and the customer satisfaction index. Record the correlation co-efficient.
- (c) Draw the least-squares line on the axes on the previous page.
- (d) Predict the customer satisfaction index for day 11.
- (e) Explain why a prediction for the customer satisfaction index for day 15 should be treated with caution.
- (f) Determine what percentage of the variation in CS Index (s) can be explained by the variation in Day (d).
- (g) Determine the average decrease in CS Index (s) per day?

Question 7**[11 marks: 4, 2, 3, 2]**

In a recent study of artists who asked for a piece of their work to be included in an exhibition, each artist was classified by the variables (i) the state they worked in and (ii) whether their piece of work was accepted by the judges.

The table below shows the number of artists in each category.

	State	NSW	VIC	QLD	WA	Total
Work accepted?	Yes	8	27		8	64
	No		86	143	39	
	Total					440

- (a) Complete the missing values and totals in the table above.
- (b) To identify the presence of an association between these two variables, explain why the state the artist worked in should be used as the explanatory variable.

Question 7 cont.

- (c) Rounding percentages to the nearest whole number, complete the percentaged two-way table below so that it may be used to identify the presence of an association between the categorical variables.

	State	NSW	VIC	QLD	WA
Work accepted?	Yes				
	No	93%			

- (d) Comment on the presence of an association between the two variables.