Baldivis
Secondary College

BALDIVIS SECONDARY COLLEGE
APPLICATIONS - Unit 3 & 4
2022 Test 1 - Bivariate Data

Student Name Answers Teacher Name Yes

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

Section 1: 20 minutes + 2 minutes reading time
Section 2: 30 minutes + 3 minutes reading time

Materials required:

Section 1 Resource free section (19 marks)
Standard writing equipment
SCSA Formula Sheet

Section 2 Calculator assumed section (30 marks)
Calculator (to be supplied by the student)
SCSA formula Sheet
One page A4 (single sided) hand written notes

Other materials allowed: Drawing templates

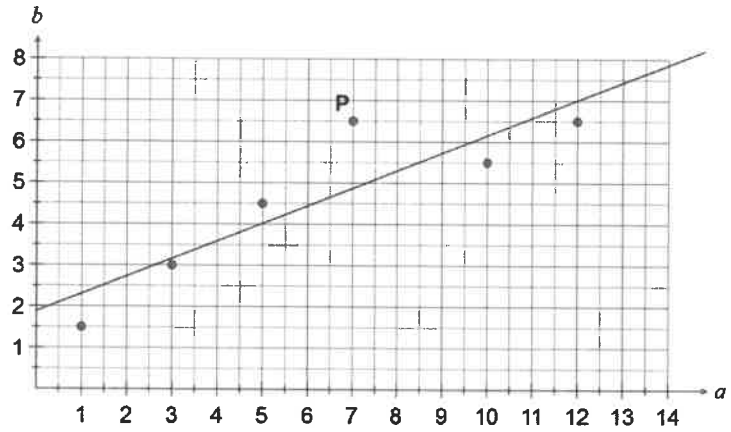
Marks available: **49 marks**

Task Weighting: **7%**

Section 1: Resource Free **[19 marks]**

Question 1. **(1,2,3,2: 8 marks)**

The scatterplot and least-squares line for a set of bivariate data (a, b) with correlation coefficient 0.85 is shown.



a) Describe the effect on the correlation coefficient if the point labelled P was removed from the dataset.

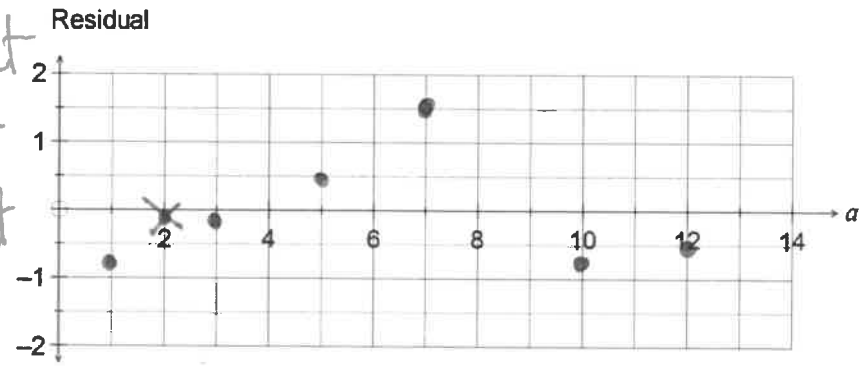
increase, move closer to 1 ✓

(b) Describe the effect on the least-squares line if all data points with $a < 4$ were removed from the dataset.

✓ - gradient decreases
✓ - vertical intercept would increase

(c) Construct a residual plot for the six paired values on the axes below.

✓ 2 points correct
✓ 4 points correct
✓ all points correct



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

Not appropriate ✓, as a pattern is evident ✓

Question 2. **(1,2: 3 marks)**

The table below shows some information about car accidents in Australia. It shows the percentage of car accidents by type of road being driven on, for three different age groups.

Type of road	Age of driver in car accident		
	18 to 30 years	31 to 50 years	Over 50 years
Inner city	43%	51%	17%
Country	32%	32%	48%
Freeway	25%	17%	35%

a) Of the car accidents involving drivers over 50 years of age, what percentage occurred on country roads or freeways?

48 + 35

= 83% ✓

b) Does the information in the table support the opinion that the age of drivers in car accidents is associated with the type of road on which they were driving? Justify your answer by quoting appropriate percentages from one age group only.

✓ Yes

✓ using %

51% of 31 to 50 have accidents on inner city compared to 32% on country and 17% on Freeway

or

48% of over 50 on countryroads

Question 3. **(2 marks)**

For the following sets of variables, state which is the explanatory variable and which is the response variable.

a) Amount of daily exercise and fitness level

Fitness - response ✓

Exercise - explanatory

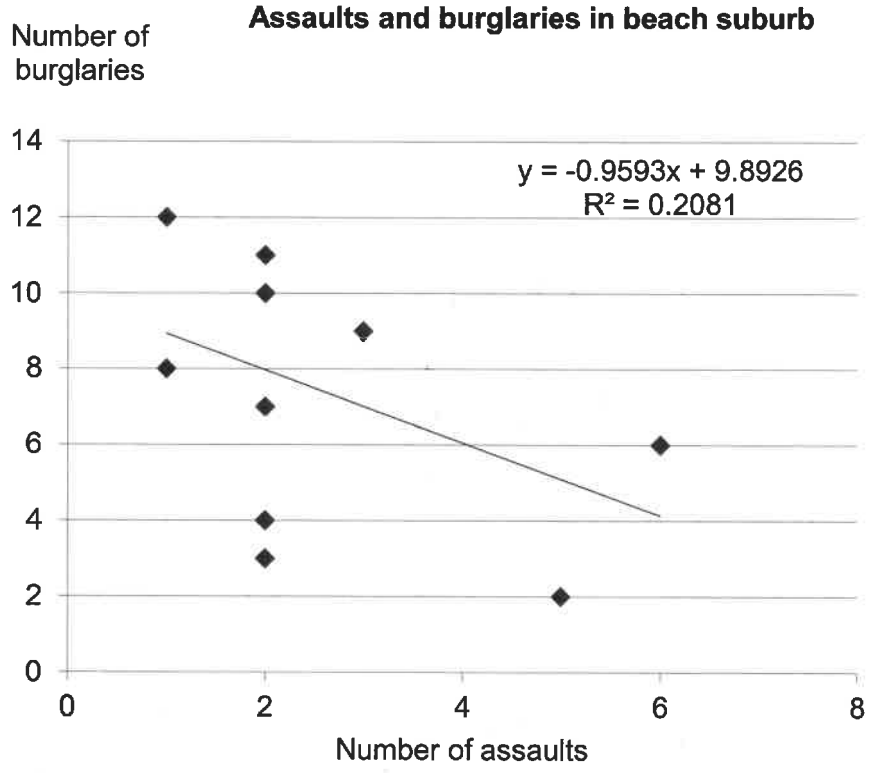
b) Price for which a car sells and the age of the car

Price - response ✓

age - explanatory

Question 4 **(2,1,2 = 6marks)**

After reading in the paper about the increased rate of criminal activity in beach suburbs, Freda decided to investigate some statistics for a popular beach suburb. She used data from the website for the Western Australia Police and examined the relationship between the number of assaults and the number of burglaries each month. Using a spreadsheet package she produced the following display.



(a) Describe the strength of the linear relationship between the two variables. Justify your conclusion. Weak ✓, data point do not form a close linear pattern ✓ or correlation coefficient is 0.43

(b) The linear relationship between the two variables is described by the equation provided.

(i) What does this equation indicate is the approximate number of burglaries in a given month when there are no assaults? ≈ 10 burglaries. ✓

(ii) What does this equation indicate would be the rate of change in the number of burglaries as the number of assaults increase?

- change is negative ✓
- rate of change is 1 ✓

As the number of assaults increases by 1, the number of burglaries decreases by 1

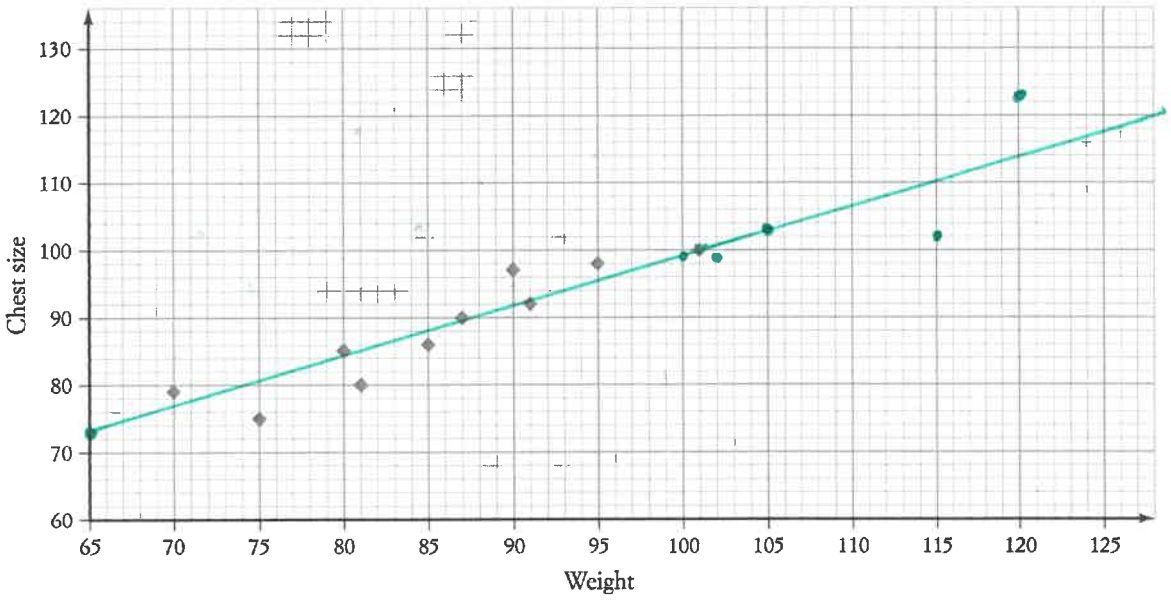
Section 2: Resource Allowed **30 marks**

Question 5 **(2, 2, 2, 2, 2, 1: 11 marks)**

Data was collected from a rugby team to investigate whether there is a relationship between a player's weight (kg) and their chest size (cm). The results are displayed in the table below.

Weight	70	75	80	81	85	87	90	91	95	101	**102**	**105**	**115**	**120**
Chest size	79	75	85	80	86	90	97	92	98	100	**99**	**103**	**102**	**123**

a) Add the last four data points from the table (**in bold**) to the scatter plot below.



b) Calculate the least-squares regression line that models the data, stating all coefficients correct to two decimal places.

~~Chest size = 0.7978x +~~

Chest size = 0.7978 × Weight + 19.588 ✓ ✓

c) Hence, sketch the regression line onto the scatter plot in part **a**.

✓ - y int (at Weight 65)
✓ - line drawn correctly

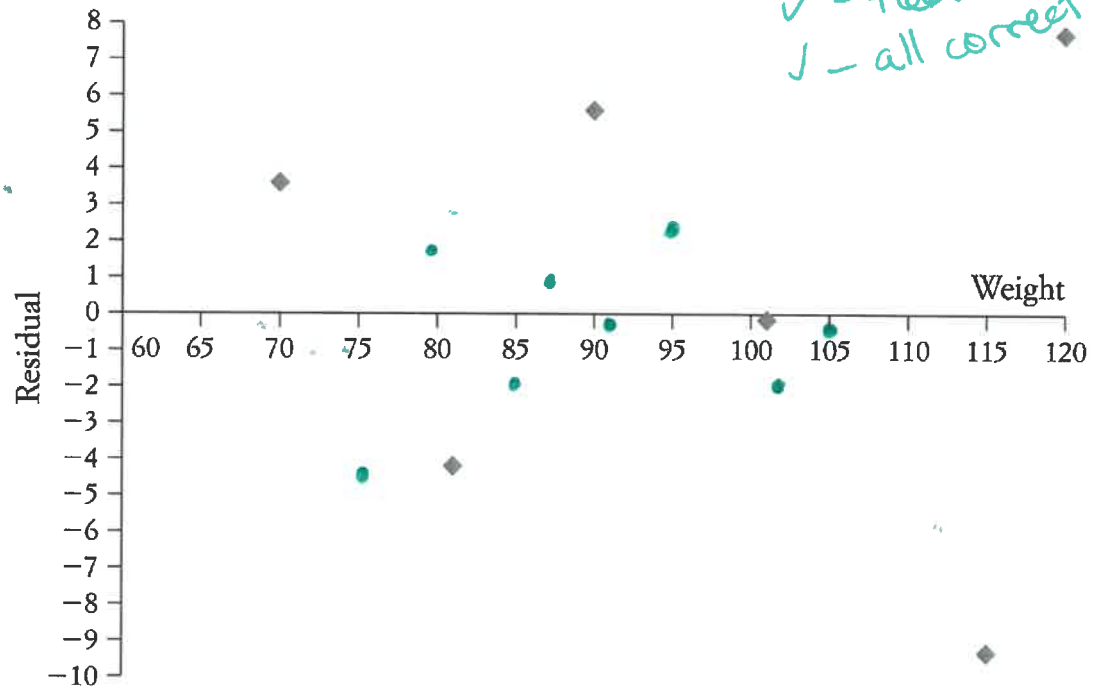
d) Calculate all the residual values (to one decimal place) for the data and complete the table below.

✓ – 4 correct
✓ – all correct

Weight	70	75	80	81	85	87	90	91	95	101	102	105	115	120
Residual	3.6	-4.4	1.6	-4.2	-1.4	1.0	5.6	-0.2	2.6	-0.2	-2.0	-0.4	-9.3	7.7

e) Hence, complete the residual plot below.

✓ – 4 correct
✓ – all correct



f) Using the residual plot, justify whether the regression model found in part **b** is appropriate.

Yes, no clear pattern.
linear mode appropriate ✓

has to given answer with reason for mark

Question 6 **[4,1,1,4 =10 marks]**

The table below represents the results of a survey that determined the age of the survey participants and whether or not they were blood donors.

(a) Complete the two-way table:

	Blood Donor	Non-donor	
16 – 18 years	8	17	25
19 – 25 years	17	42	59
26 – 40 years	29	51	80
41 – 65 years	9	33	42
	63	143	206

✓ - correct row totals
✓ - correct column totals
✓ - 4 entries correct
✓ - all correct.

(b) How many 26–40 year olds were surveyed?

80 ✓

(c) How many of the participants surveyed were blood donors?

63

(d) Convert the two-way table to a percentage two-way table.

	Blood Donor	Non-donor	
16 – 18 years	32%.	68%.	100%.
19 – 25 years	29%.	71%.	100%.
26 – 40 years	36%.	64%.	100%.
41 – 65 years	21%.	79%.	100%.

✓ - row %
✓ - 2 rows correct
✓ - all rows correct
✓ rounding

✓ - each correct row

if did column %.
2 marks if correct.

Question 7 **(2,1,2 = 5marks)**

The least squares regression line between M and s is given by $M = 0.07s + 3.24$.

a) Find the average increase in M corresponding to an increase of 30 units in s.

gradient = 0.07 $0.07 \times 30 = 2.1$ ✓ ✓

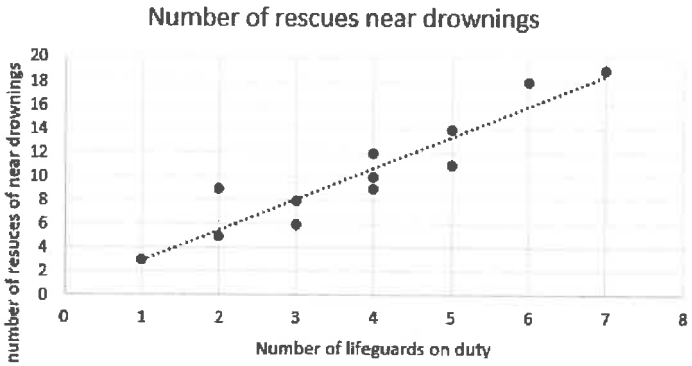
b) Predict the value of M when s=60

$M = 0.07(60) + 3.24$
$= 7.44$ ✓

c) The actual value of M when s=60 is 6.32. Find the residual associated with the prediction in b).

$6.32 - 7.44$ ✓
$= -1.12$ ✓

Question 8 **(2, 2= 4 marks)**



(a) Comment on the relationship between the number of lifeguards on duty and the number of rescues of near drownings.

✓ – relationship
As lifeguards increase, rescues increase ✓ – answer in context

(b) Jodie calculated the value of the correlation coefficient for the given data to be 0.96 (to 2.d.p.) and concluded that the more lifeguards were on duty, the more near drowning rescues occurred and hence it was probably better to reduce the number of lifeguards on duty. Discuss Jodie's conclusions.

✓ – correlation ≠ causation
✓ – reasonable external variable
n° swimmers, etc.