

BALDIVIS SECONDARY COLLEGE
APPLICATIONS - Unit 3 & 4
2020 Test 1 - Bivariate Data

Student Name Answer Key Teacher Name _____

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

Section 1: 17 minutes + 2 minutes reading time

Section 2: 33 minutes + 3 minutes reading time

Materials required:	Section 1 Resource free section	(17 marks)
	Standard writing equipment SCSA Formula Sheet	
	Section 2 Calculator assumed section	(33 marks)
	Calculator (to be supplied by the student) SCSA formula Sheet One page A4 (single sided) hand written notes	

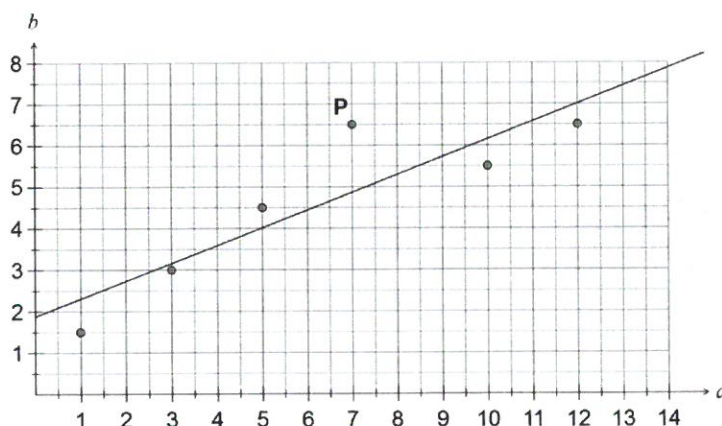
Other materials allowed: Drawing templates

Marks available: 50 marks

Task Weighting: 7%

Question 1. [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. (1 mark)

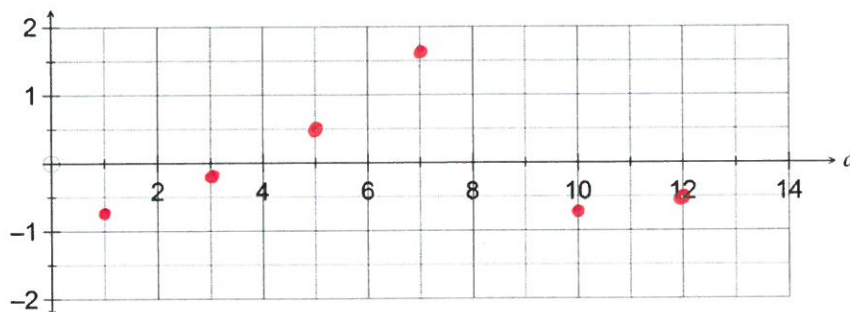
Correlation coefficient would increase, becoming closer to 1.

- b) Describe the effect on the least-squares line if all data points with $a < 4$ were removed from the dataset. (2 marks)

- gradient would increase ✓
- vertical intercept would decrease ✓

- c) Construct a residual plot for the six paired values on the axes below. (3 marks)

Residual



✓ - 2 points plotted correctly
✓ - 4 points plotted correctly
✓ - All points plotted correctly.

- d) Comment, with reasons, on the appropriateness of fitting a linear model to this dataset. (2 marks)

Not appropriate, as a pattern is evident with the points in the residual plot. ✓

Question 2. [9 marks]

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
Male under 20 years	8	22	6
Female under 20 years	6	20	8
Male 20 to 25 years	26	7	3
Female 20 to 25 years	30	9	5
Male over 25 years	24	3	2
Female over 25 years	18	2	1

a) Complete the two-way table below.

(2 marks)

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

✓ - at least 3 correct entries
✓ - all correct.

b) The incomplete table below shows row percentages. Complete the table.

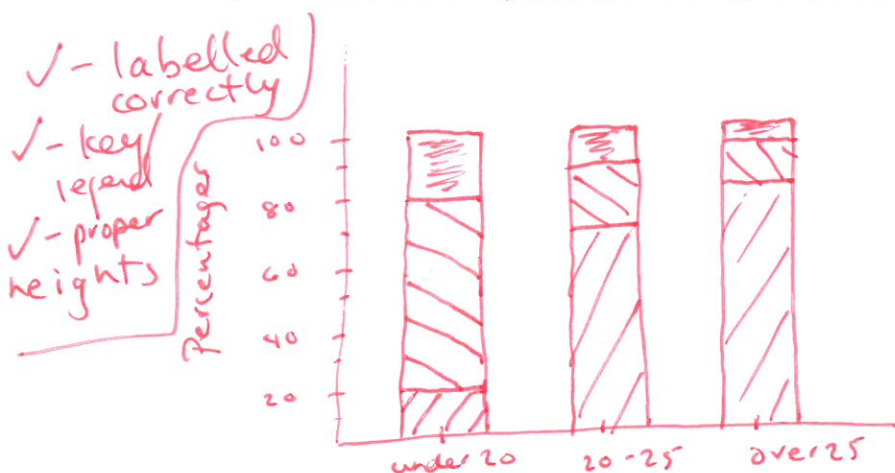
(2 marks)

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

✓ - at least 3 correct
✓ - all correct.

c) Use the two-way table to draw a segmented column graph

(3 marks)



Key: - Agree
 - Disagree
 - Undecided

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

(2 marks)

✓ - link stating association
✓ - link gives reasoning

As age increases, the % of students who agree increases. Percentages in the agree column increase with age (20% vs. 70% vs. 84%) or some other statistical comparison.

Answers

37

Section 2: Calculator Assumed [33 marks]

Question 3. [9 marks]

The table below shows the paired scores of twelve students in a Chemistry and a Physics exam, and the Chemistry score of a thirteenth student who was absent for the Physics exam.

Student	Chemistry score (C)	Physics score (P)
1	53	40
2	63	46
3	95	70
4	66	42
5	67	47
6	63	55
7	70	59
8	70	63
9	80	58
10	77	61
11	74	67
12	80	66
13	84	-

- a) Name the explanatory variable. (1 mark)

Chemistry score

- b) Calculate the correlation coefficient r_{cp} . (1 mark)

0.812

- c) Graph the data on your calculator and use features of the graph to explain why the Physics teacher thought that it would be reasonable to predict the Physics score of Student 13 from their Chemistry score. (2 marks)

The points display a straight line pattern, showing a moderate to strong positive linear association.

- (d) Determine what percentage of the variation in the Physics scores can be explained by the variation in the Chemistry scores. (2 marks)

$$r^2 = (0.812)^2 = 0.659 \quad \checkmark 1 \text{mk.}$$

66% of the variation

- (e) Determine the equation of the least-squares line that models the relationship between the Chemistry and Physics scores, rounding coefficients to two decimal places. (2 marks)

$$P = 0.78C + 1.32$$

- 1mk if x & y used.
- 1mk if not to 2dp.

- (f) Predict the Physics score of Student 13. (1 mark)

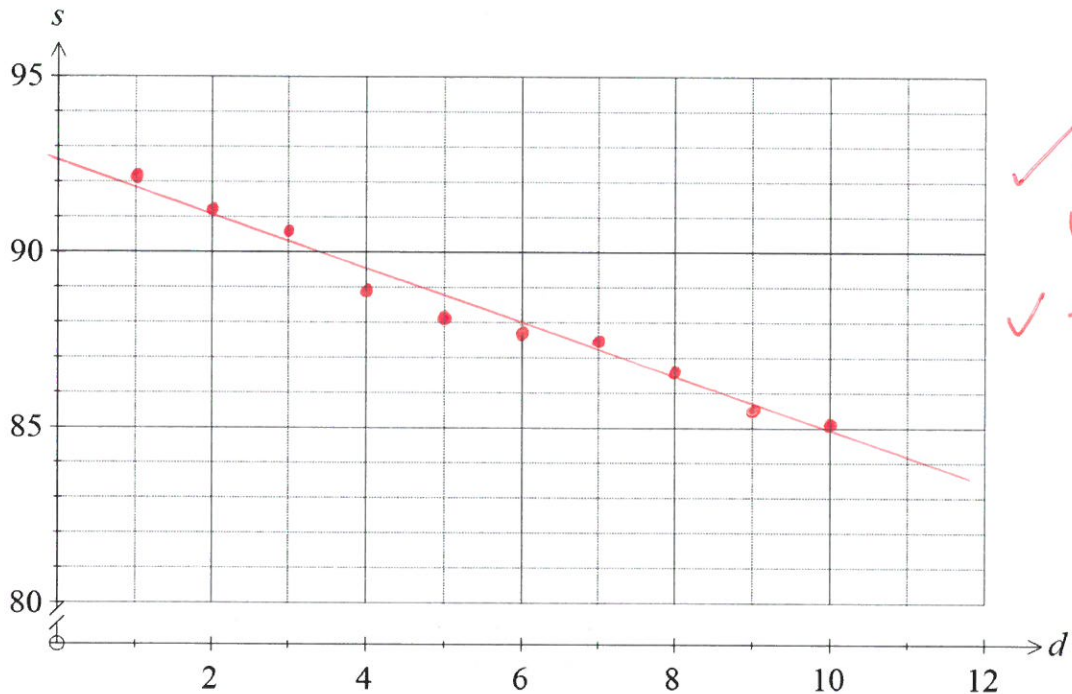
$$P = 0.78(84) + 1.32 = 66.84 \approx 67$$

12
Question 4 [8 marks]

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. (2)



✓ - 3 or less errors
✓ - No errors.

- b) Determine the equation of the least-squares regression line that models the linear relationship between the explanatory and response variables. (2 marks)

$$s = -0.78d + 92.59$$
 - 1mk if x & y used
 - 1mk if less than 2 dp.
- c) Draw the least-squares line on the axes above. (2 marks)
 see above.
- d) Calculate and interpret Pearson's correlation coefficient for the data. (1 mark)
 $r = -0.988$ strong negative linear correlation.
- e) Calculate and interpret the coefficient of determination for the data. (1 mark)
 $r^2 = 0.976 \approx 98\%$ of the variation in customer satisfaction index explained by the day.
- f) Is the regression line found in part b) an appropriate linear model? Justify your answer. (2 marks)
 ✓ Yes it would seem so, based on high r -value (-0.988).

- g) Predict the customer satisfaction index for day 11.

(1 mark)

$$84.02 \approx 84$$

- h) Explain why a prediction for the customer satisfaction index for day 15 should be treated with caution.

(1 mark)

Extrapolation involved, so despite high negative r -value, view with caution.

Question 5 [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:

(4 marks)

	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 entries correct.

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

(1 mark)

80

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

(1 mark)

63

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

(4 marks)

	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 percent
 ✓ - 2 rows correct
 ✓ - all rows correct
 ✓ - correct rounding.

Question 6 [6 marks]

Roe (eggs) collected from Salmon are considered a delicacy by some people.

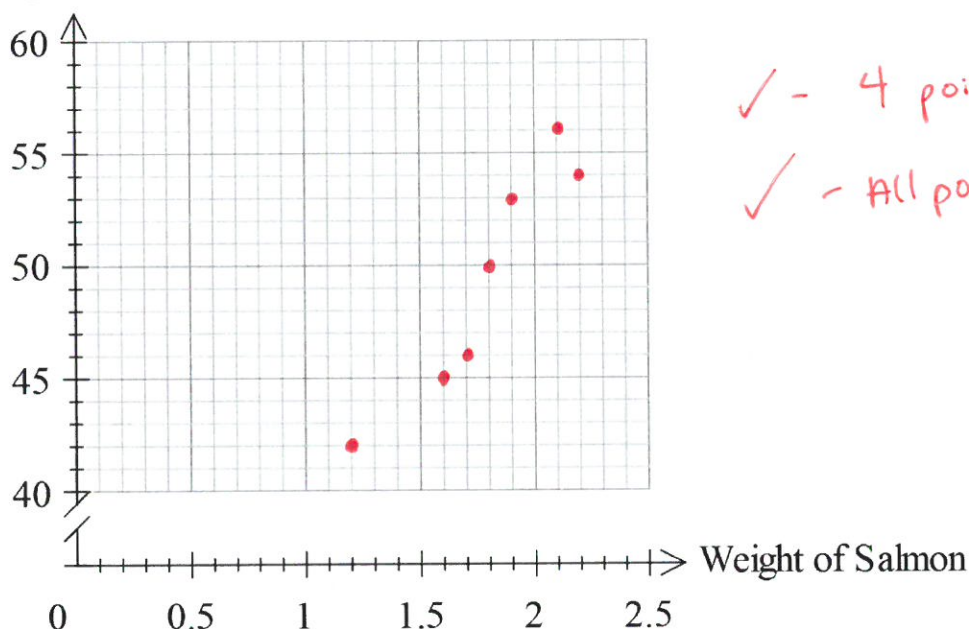
Seven salmon, all with roe, are used to analyse the relationship between weight of a salmon, in kg, and the weight of its roe, in grams. The resulting data are recorded in the table below.

X	Weight of Salmon (kg)	1.2	1.6	1.7	1.8	1.9	2.1	2.2
Y	Weight of Roe (g)	42	45	46	50	53	56	54

- a) Show the data on the axes below.

(2 marks)

Weight of Roe



✓ - 4 points correct
✓ - All points correct.

- a) You are asked to predict the weight of roe from a salmon whose weight is 750g. Comment on this request.

(2 marks)

It is extrapolation - ✓ 1mk
unlikely to be reliable - ✓ 1mk.

- b) What is the average rate of increase in the weight of roe for each kilogram increase in the weight of the salmon?

(1 mark)

14.72g per kg.

- c) The researcher concludes that it is the increase in weight of the salmon which causes the increase in the weight of roe. Comment on this claim.

(1 mark)

Correlation does not mean causality.