Mathematics Department

Course: ATMAA

Topic Title: Bivariate data

Test 2

Student Name: Solutions

Date: _____

Special Instructions: **Calculator Allowed**

1 page of A4 notes and Formula Sheet Allowed

Time Allowed: 60 mins

Marks: / 58

Question 1. (3, 2, 2, 5, 2: 14 marks)

315 students were interviewed about their favourite subject – Mathematics or English. Of the 170 girls, 70 preferred Mathematics, while 45 boys preferred English.

a) Construct a two way table showing this information.

	Maths	English	
Girls	70	100	170 ✓
Boys	100	45	145 ✓
	170	145	315 ✓

b) Determine the explanatory and response variable.

Favourite Subject - RV ✓

Boys / Girls - EV ✓

c) Construct an appropriate percentage two way table.

	Maths	English	
Girls	41%	59%	100% ✓
Boys.	69%	31%	100%. ✓

Maths
Eng

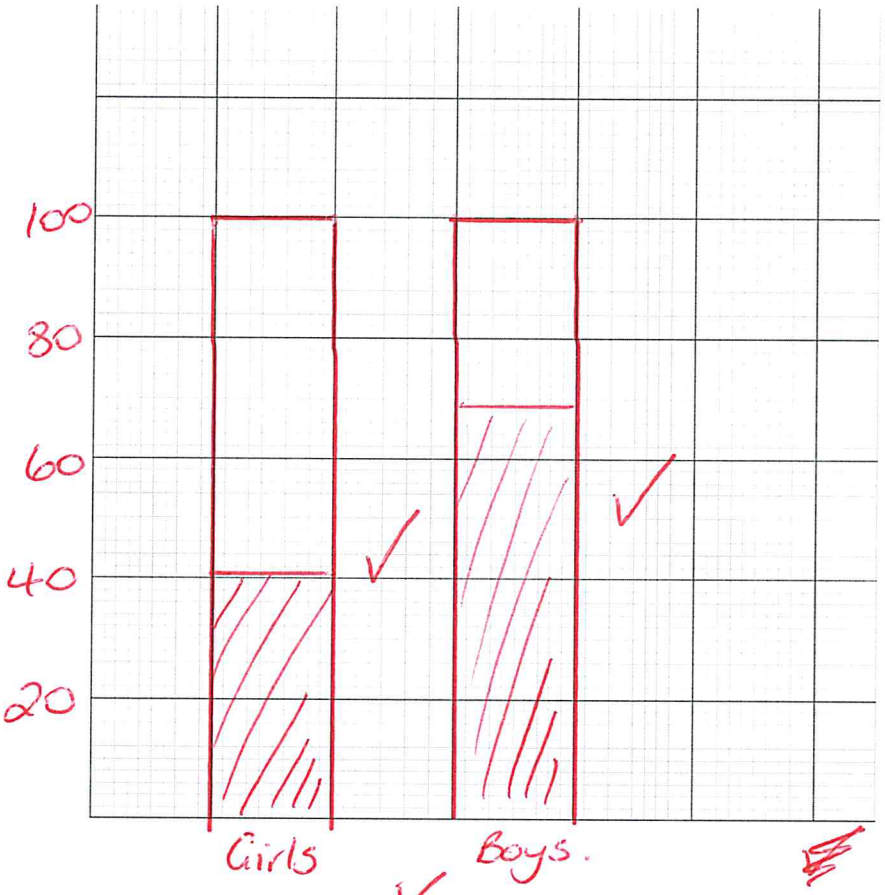
Girls
41%
59%
100%

or

Boys.
69%
31%
100%.

7

d) Construct a segmented column graph.



✓ Scale ✓ ✓ ✓

e) Is there an association between the two variables? Explain your answer.

Yes, higher % of boys prefer Maths (69%) and a higher % (59) of girls prefer English ✓

Suggests that there maybe an association with prefered subject and gender. ✓

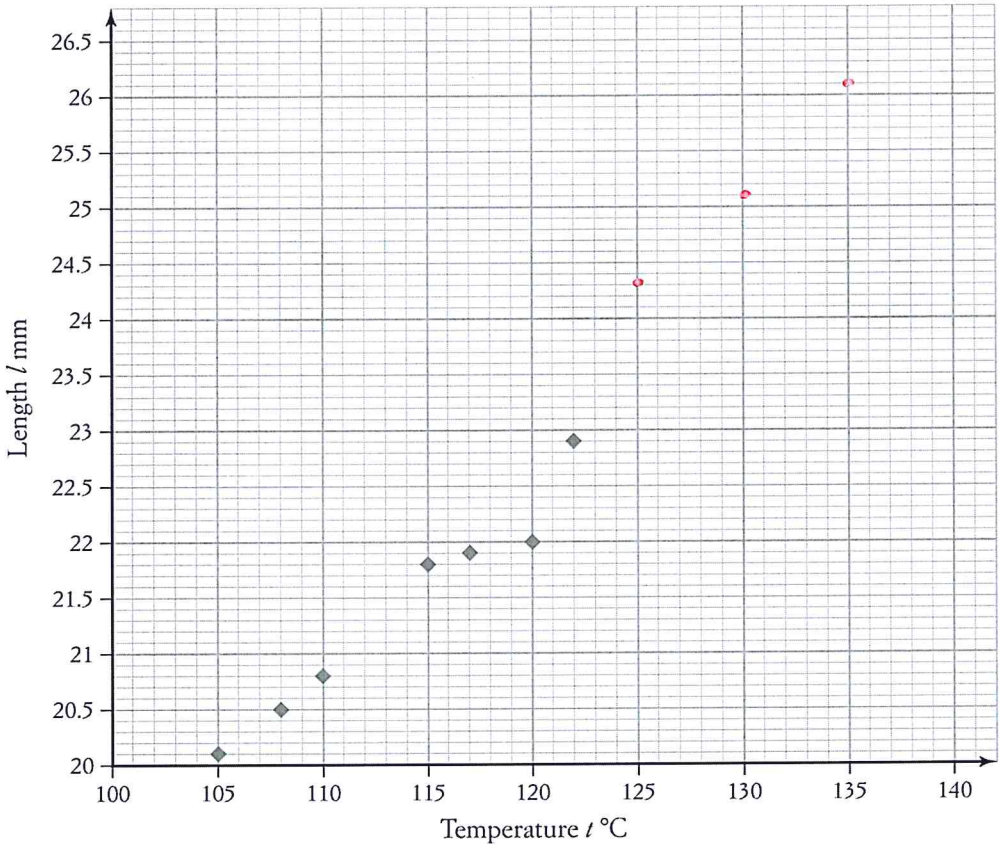
7

Question 2. (1, 1, 2: 4 marks)

The length (l mm) of a metal bar was measured at various temperatures (t°C) to give the following results.

t°C	105	108	110	115	117	120	122	**125**	**130**	**135**
l mm	20.1	20.5	20.8	21.8	21.9	22.0	22.9	**24.3**	**25.1**	**26.1**

a) Complete the following scatter plot by adding the last three points shown in bold in the table.



b) Calculate Pearson's correlation coefficient for the data, correct to two decimal places.

$r = 0.98$ ✓

c) Interpret the relationship between the variables 'Length' and 'Temperature', referring to form, direction and strength.

A positive ✓ strong linear ✓ relationship.

Question 3. (3, 2, 2, 2: 9 marks)

Data was collected to investigate whether a person's income ($) depends on their height (cm) and is displayed in the table below. State all answers correct to two decimal places.

Height (cm)	174	162	165	183	185	158	176	196
Income ($)	62 000	36 000	44 000	51 000	60 000	47 000	39 000	26 000

a) Using CAS, calculate the least-squares regression line that models the data.

$a = -111.92$ $b = 65197.34$. ✓

$y = -111.92x + 65197.34$.

✓ ✓

b) Calculate and interpret Pearson's correlation coefficient for the data.

$r = -0.12$. ✓

no relationship. ✓

c) Calculate and interpret the coefficient of determination for the data.

$r^2 = 0.014$ ✓

1% of the variation can be explained.

or

99% of the variation can not be explained

there not linear. ✓

d) Is the regression line found in part **a** an appropriate linear model? Justify your answer.

NO, the r value = −0.12 ✓ and the $r^2 = 0.01$.

Show that there is no linear relationship.

✓

The following information relates to Questions 4 and 5.

A group of 25 year olds were surveyed regarding the number of hours spent exercising per week and their resting heart rate (beats per minute). The data collected for the number of hours of exercise per week ranged from 0 hours to 6 hours.

The regression line that models the data was found to be:

$$\text{Resting heart rate} = 63 - 0.96 \times \text{Exercise},$$

with a correlation coefficient of –0.52.

Question 4. (7 marks)

Use the information above to complete the following sentences.

a) A person's resting heart rate will be 63 beats per minute when the amount of exercise is 0 hours per week.

b) 27% of the variation in Resting Heart Rate can be explained by the variation in the amount of exercise.

$-0.52 = r$.

$r^2 = 0.27$.

c) A person's resting heart rate will decrease by 0.96 beats per minute for every one hour increase in exercise.

Question 5. (2, 2, 2: 6 marks)

Use the regression line to:

a) Predict the resting heart rate for a person who exercises for 3 hours per week. Is this prediction classified as interpolation or extrapolation? State answer correct to one decimal place.

$y = 63 - 0.96x$

$y = 63 - 0.96(3)$

$y = 60.1$

interpolation.

b) Predict the amount of exercise required per week for a person with a resting heart rate of 55 beats per minute. Is this prediction classified as interpolation or extrapolation? State answer correct to one decimal place.

$55 = 63 - 0.96x$

$x = 8.3$ hours.

extrapolation.

c) Which of the above predictions is the most reliable? Justify your answer.

Prediction a) is the most reliable. as the amount of hours is within the range of 0-6 hours.

Question 6. (3, 2, 2: 7 marks)

The weights (kg) of 13-year-old students (y) and the number of chocolate bars consumed per week (x) were recorded and the values of the following statistics were determined.

$\bar{x} = 3.23$ $s_x = 1.6$ $\bar{y} = 41.55$ $s_y = 5.61$ $r = 0.56$

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

$a = \frac{r \times s_y}{s_x}$

$= 0.56 \times \frac{5.61}{1.6}$

$a = 1.96$ ✓

$b = \bar{y} - a\bar{x}$

$= 41.55 - 1.96(3.23)$

$b = 35.22$ ✓

$y = 1.96x + 35.22$ ✓

b) Use the regression line to predict the weight of a 13-year-old student who eats 5 chocolate bars per week, correct to two decimal places.

$y = 1.96(5) + 35.22$ ✓

$= 45.02$ ✓

c) Hence, calculate the residual value for Peter who weighs 43 kg and eats 5 chocolate bars per week.

Res. $= y - \hat{y}$

$= 43 - 45.02$ ✓

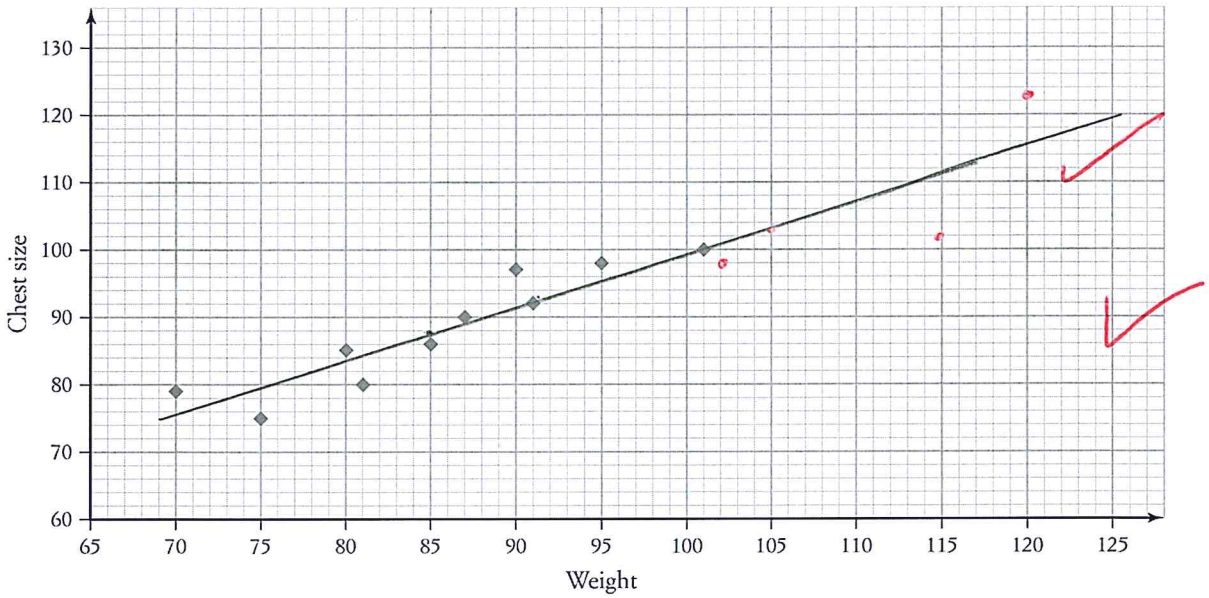
$= -2.02$ ✓

Question 7. (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.

CAS

$a = 0.80$

$b = 19.59$.

$y = 0.8x + 19.59$. ✓

✓ ✓

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

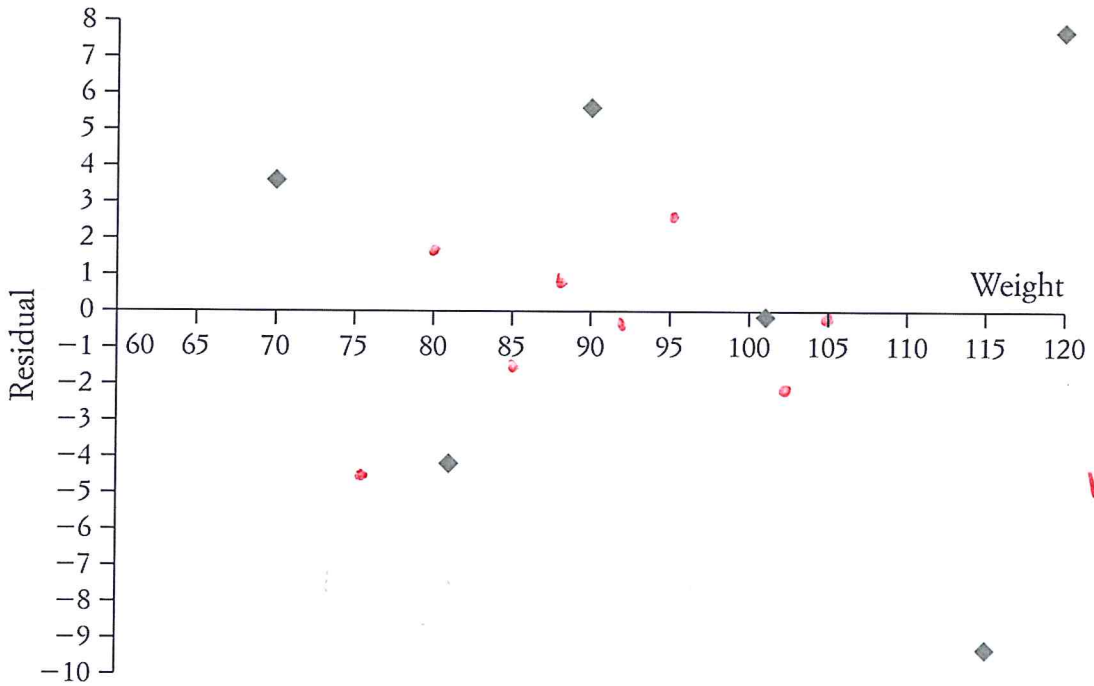
✓✓

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

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	-0.4	-9.3	7.7

✓✓

e) Hence, complete the residual plot below.



✓✓

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

Regression model $y = 0.80x + 19.59$. is possibly appropriate, since the residual plots are scattered above and below 0.

✓