Name…………………………………………………………………………………………………...

No calculators nor notes allowed.   Formula sheet allowed.

27 mark total.   25 minute time limit.

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

Members of a gym were asked if they did Cardio or Weight training. Each responder did one kind of training. The table below summarises the results.

	Cardio	Weights
Male	12	36
Female	24	6

(a)   Determine the missing percentages.

	Cardio	Weights
Male		
Female	80%	20%

(b)   Compare the percentages of males and females who train at the gym using *weights*.
       Comment on your results in regard to male and female preferences.

Question 2 [3 marks: 1, 2]

The scatterplot below shows the wrist circumference and ankle circumference, both in centimetres, of 13 people. A least-squares regression line has been fitted to the scatterplot with ankle circumference as the explanatory variable.
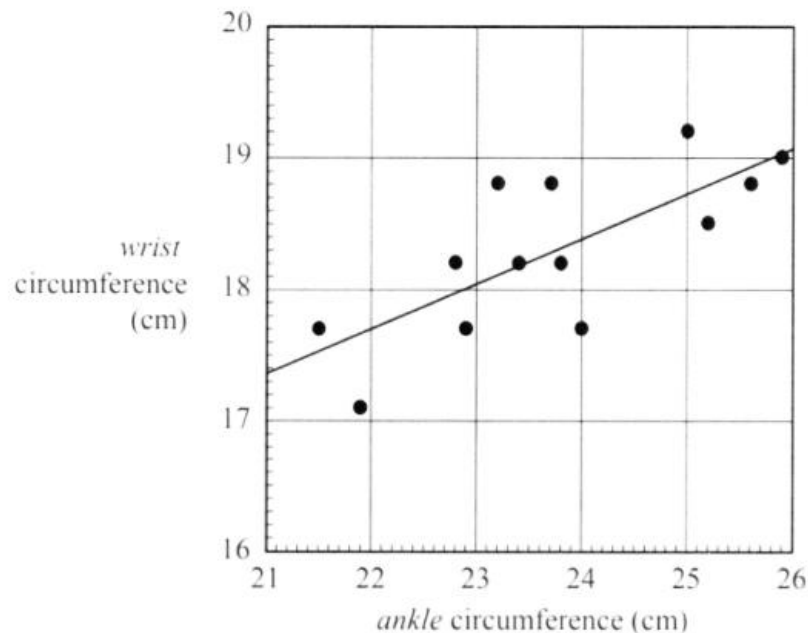
(a) Which of the following equations is the closest match to the least-squares regression line?

A.   ankle = 10.2 + 0.342 x wrist

B.   wrist = 10.2 + 0.342 x ankle

C.   wrist = 17.4 + 0.342 x ankle

D.   wrist = 17.4 + 0.731 x ankle

(b) Use the least-squares regression line selected in part (a) above to predict the wrist circumference of the person with an ankle circumference of 24 cm. Clearly show how this prediction could be derived from the graph.

Question 3 [6 marks: 2, 2, 2]

A study was conducted to investigate the relationship between the number of moths caught in a moth trap and the trap type. The moth count in each trap was classified as one of 'More than 500', 'Between 250 and 500' and 'Less than 250'. The results were summarised in the percentaged segmented bar chart below.
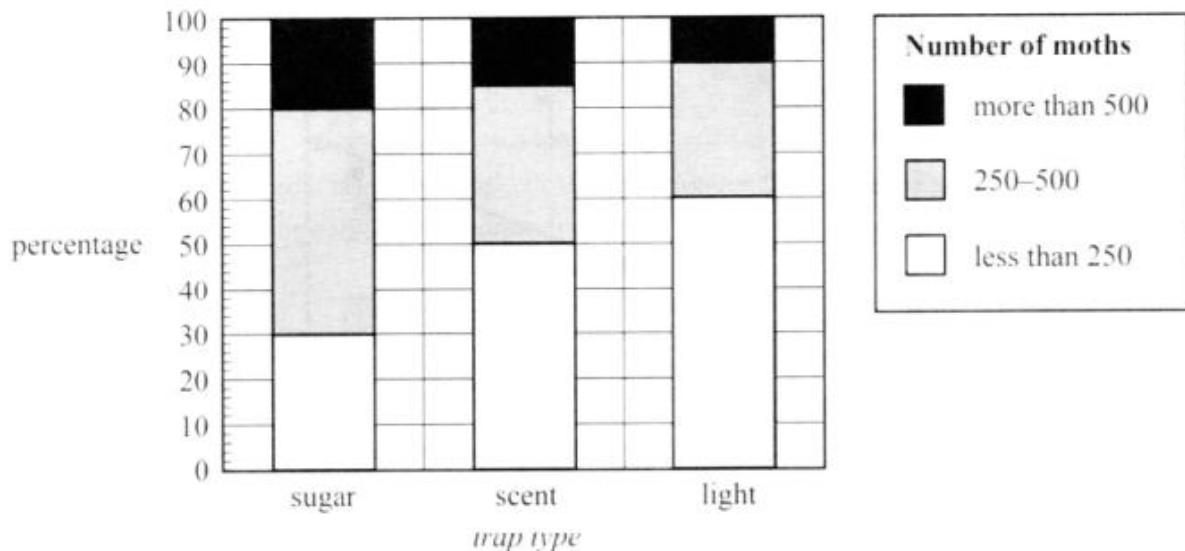
(a) There were 300 light traps used in the study. Find the number of light traps that caught fewer than 250 moths.

(b) In the category '250-500', there are 200 sugar traps.

Find the number of sugar traps in the study.

(c) Comment on the association between the trap type and the number of moths caught in a trap.

Question 4 [6 marks: 2, 1, 3]

A study was conducted to investigate the association between the number of female moths (m) caught in a trap set in a forest and the moth eggs per square metre (e) in the forest. Both the moth and the eggs are from the same species.
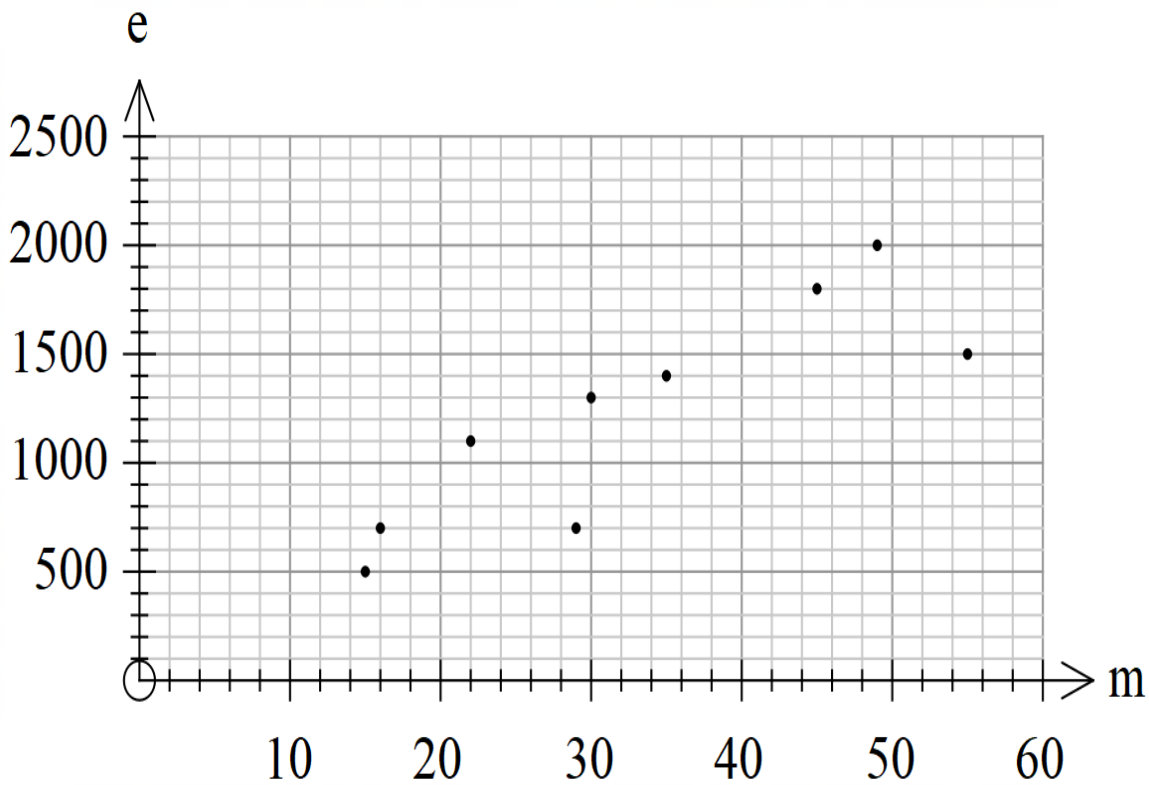
The equation of the least-squares regression line is $e = 195 + 31 \times m$

(a) Draw the least-squares regression line on the scatterplot above. Show your calculations clearly.

(b) Interpret the slope of the regression line in terms of the variables egg density and number of female moths caught in the trap.

Question 4 cont.

(c)   Describe the association between the egg density and number of female moths caught in the traps in terms of strength, form and direction.

Question 5   **[2 marks]**

In a recent survey it was found there was a high correlation between the amount of alcohol consumed in a household and the number of motor vehicle accident claims from members of that household.

Does this mean that alcohol consumption causes accidents? Explain your answer.

Question 6 [6 marks: 2, 2, 2]

Three groups of bivariate data were analysed. The correlation coefficients and/or the coefficients of determination were determined.

(a) The correlation coefficient for Group A was $r = 0.87$

    (i) Describe the association that exists between the data set in this group.

    (ii) An outlier was included in the calculation of the correlation coefficient of 0.87. What would you expect would happen to the correlation coefficient if the outlier was removed from the calculation? Justify your answer.

(b) Determine the difference in the correlation coefficient values of Groups B and C if it was discovered that for Group B, $r = 0.8$ and for Group C, $r^2 = 0.25$. Both groups have a positive correlation coefficient.

	Greenwood College Year 12 Applications Test 3 Bivariate Data 2020 Resource-Allowed

Name……………………………………………………………………………………………...

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

23 mark total.   25 minute time limit.

Question 7   **[6 marks: 3, 3]**

A survey was done at a local University to determine if there was an association between the student's year of study and their choice of living arrangement.

The table below indicates the response of the 261 students interviewed.

	1st Year	2nd Year	3rd Year
Parents Home	35	26	15
On Campus	43	40	21
Private Accommodation	15	20	46

(a)   Recreate the table showing either row or column percentages as appropriate. Round to whole numbers.

	1st Year	2nd Year	3rd Year
Parents Home			
On Campus			
Private Accommodation			

(b)   Comment on the association between the two variable, explaining your reasoning.

Question 8 [9 marks: 2, 3, 2, 2]

A study was done to in a series of investigate the possible association between the number of hours that a student exercises per week and the time (measured in minutes) that the student takes to complete a series of physical tasks. Based on the data in the table below, the equation for the regression line is given by $\hat{y} = 3.9 + 0.7x$
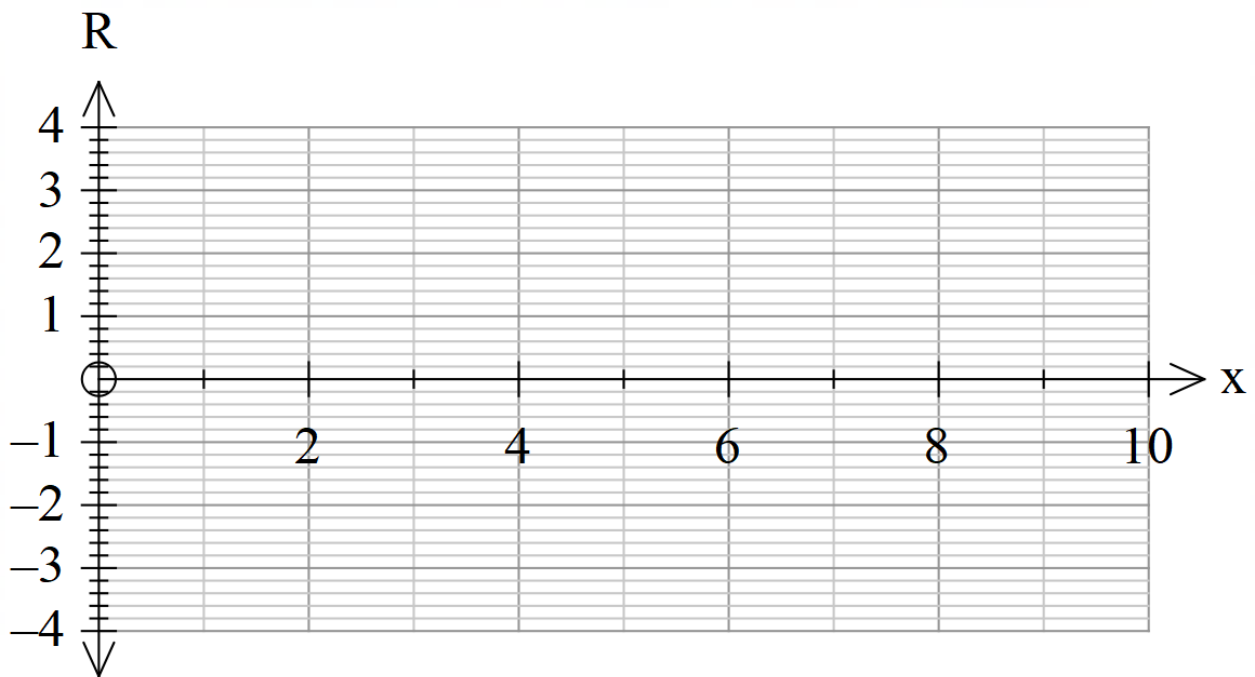
x	1	5	8	6	**C**	9	4
y	3	9	11	9	11	7	7
$\hat{y}$	4.6	7.4	**B**	8.1	10.9	10.2	6.7
residual	-1.6	1.6	1.5	**A**	0.1	-3.2	0.3

(a) Interpret the numbers 3.9 and 0.7 (from the equation of the regression line) in the context of this situation

(b) Find the values of **A**, **B** and **C**.
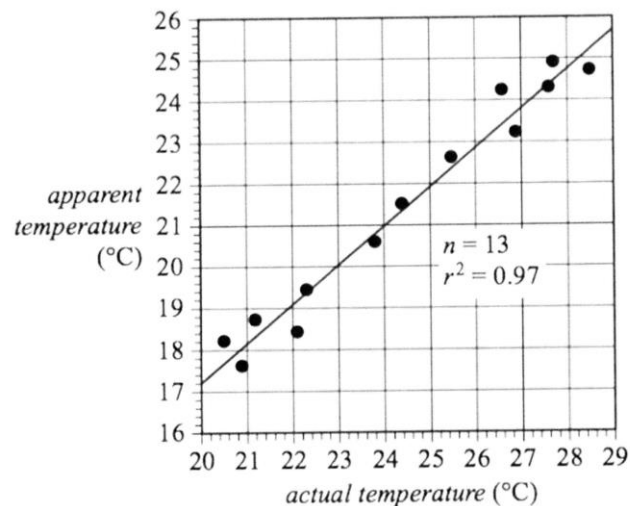
Question 8 cont.

(c)   Graph the residuals.

R

4
3
2
1

$\rightarrow$ x

−1   2   4   6   8   10
−2
−3
−4

(d)   Determine whether a linear regression model is then appropriate model to be used. State the reason for your answer.

Question 9   **[8 marks: 1, 1, 1, 2, 1, 2]**

The table below shows a sample of actual temperatures and apparent temperatures recorded at the weather station. A scatterplot of the data is also shown below right.

The data will be used to investigate the association between the two variables; *apparent temperature* and *actual temperature*.

Apparent temperature (°C)	Actual temperature (°C)
24.7	28.5
24.3	27.6
24.9	27.7
23.2	26.9
24.2	26.6
22.6	25.5
21.5	24.4
20.6	23.8
19.4	22.3
18.4	22.1
17.6	20.9
18.7	21.2
18.2	20.5

(a) State the explanatory variable.

(b) The co-efficient of determination for the association between the variables *apparent temperature* and *actual temperature* is 0.97. Interpret the coefficient of determination in terms of these variables.

Question 9 cont.

(c)   The least-squares line that can be used to predict the *apparent temperature (a)* from *actual temperature (t)* is $a = t + 2.4$

Write the values of:

   (i)   the slope of the least-squares regression line.

   (ii)   the value of the predicted apparent temperature when the actual temperature is 0°C.

(d)   The least-squares regression line was used to predict the apparent temperature on a day when the actual temperature was 40°C.

   (i)   What was the prediction for the apparent temperature?

   (ii)   Can this prediction be considered with a high degree of confidence? Comment in the context of the collected data.