Name: _____

Attached you will find revision material for:

Core Module: Data & Analysis

Further Mathematics

Theory covered:

- Displaying, summarising and describing **univariate** data (Ch 1 & 2)
- Displaying, summarising and describing relationships in **bivariate** data (Ch 3)
- Introduction to **regression** analysis (Ch 4 & 5)
- Displaying, summarising and describing **time series** data (Ch 6)

31

Due date:

You should complete all questions for this section in the spaces provided.

Question 1

Ten Grade 6 students were randomly chosen at a school and asked to estimate how many hours a week they were involved in physical activity such as cycling or basketball.

Their responses together with their mass in kg are shown in **Table 1** below.

Table 1

Student	Physical activity a week (hours)	Mass (kg)
A	7	50
B	4	55
C	12	44
D	15	41
E	10	45
F	6	49
G	9	46
H	11	43
I	5	51
J	7	48

a. Find the mean and the standard deviation of the mass of the students.

Express your answers correct to 3 significant figures.

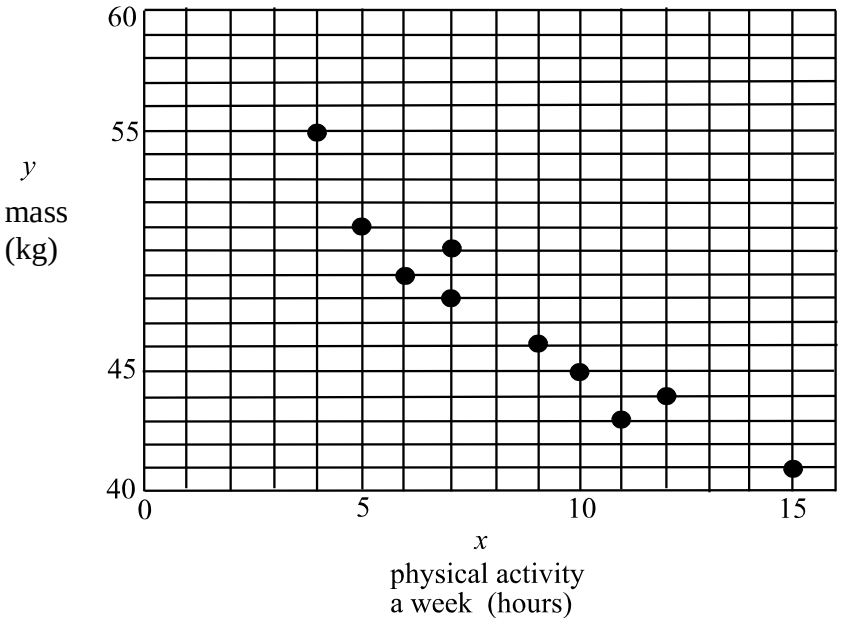
mean = ☐ kg , standard deviation = ☐ kg

2 marks

b. The interquartile range of the mass of these ten students is ☐ kg

1 mark

On the scatterplot below, the points representing the data from **Table 1** have been plotted.



c. In this analysis, the explanatory variable is _____

1 mark

d. The equation of the least squares regression is to be calculated for this set of data.

It has been assumed therefore that the relationship between the number of hours of physical activity that a student engages in and their mass, is ☐

1 mark

e. The equation of the least squares regression line is given by

mass = ☐ − ☐ × number of hours of physical activity a week

Express answers correct to two decimal places.

2 marks

f. We know that ☐ % of the variation in the mass of these students can be explained by the variation in the number of hours a week that they engage in physical activity.

Express your answer correct to the nearest whole number.

1 mark

g. How much does the equation of least squares regression overestimate or underestimate the mass of student A?

_____

2 mark

Question 2

At the same school, the numbers of Grade 5 boys and girls who engage in 10 hours or more of physical activities a week together with the number of Grade 5 boys and girls who engage in less than 10 hours a week is shown in **Table 2** below.

Table 2

	boys	girls	Total
10 hours or more a week	17	10	27
Less than 10 hours a week	6	12	18
Total	23	22	

a. Percentage **Table 2** by calculating column percentages and write your answers in **Table 3** below. Express your percentages correct to 2 decimal places.

Table 3

	boys	girls
10 hours or more a week		
Less than 10 hours a week		

2 marks

b. In percentaging **Table 2** by calculating column percentages, which two categories of grade 5 students are we comparing?

_____

1 mark

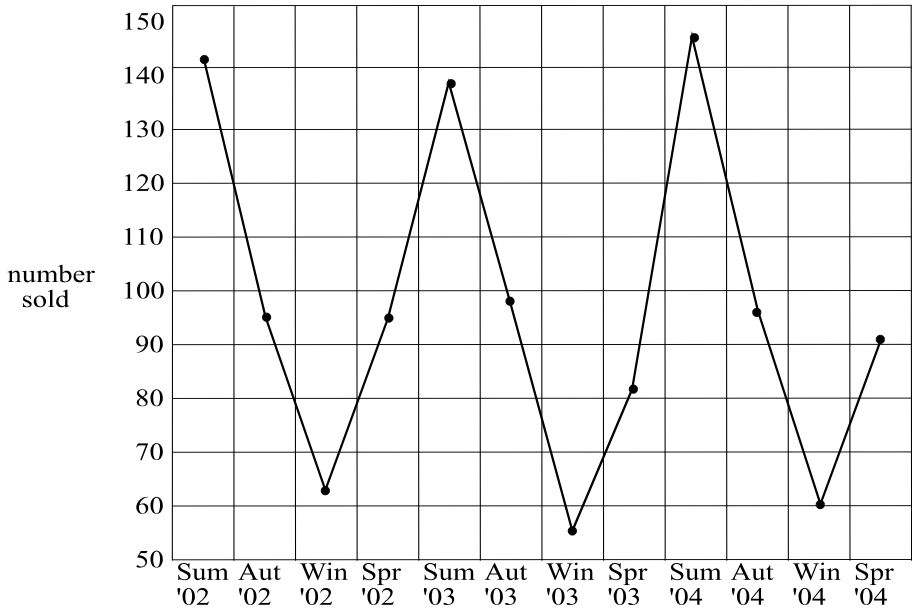
Question 3

A slimming business displays its seasonal sales in 'Slimming Plans' for 2002 – 2004 in **Table 4** below.

Table 4

Season	Sum 02	Aut 02	Win 02	Spr 02	Sum 03	Aut 03	Win 03	Spr 03	Sum 04	Aut 04	Win 04	Spr 04
Number of Slimming Plans sold	142	95	63	95	137	98	55	82	145	96	60	91

A time series plot for this data is shown on the graph below.



a. Describe the data shown on the time-series plot.

_____

1 mark

b. By using the three-median smoothing technique, smooth the time series shown on the graph above. Plot the smoothed time series on this same graph.

2 marks

Question 4

Ten Year 11 students were asked to estimate their average travelling time to and from school each day. Their responses are given in **Table 5** below.

Travelling time (mins)	5	12	18	27	45	46	50	62	70	73

Table 5

a. Calculate the mean and the standard deviation for this set of data.
Express your answers correct to two decimal places where appropriate.

mean = ☐ standard deviation = ☐

2 marks

These ten students were also asked to estimate the average amount of time they spent doing homework on a weeknight.

A scatterplot of the average daily homework time against the average daily travelling time for each of the ten students is shown in **Diagram 1** below. A least squares regression line has been fitted to the data. The scale on the average daily homework time axis has been omitted.

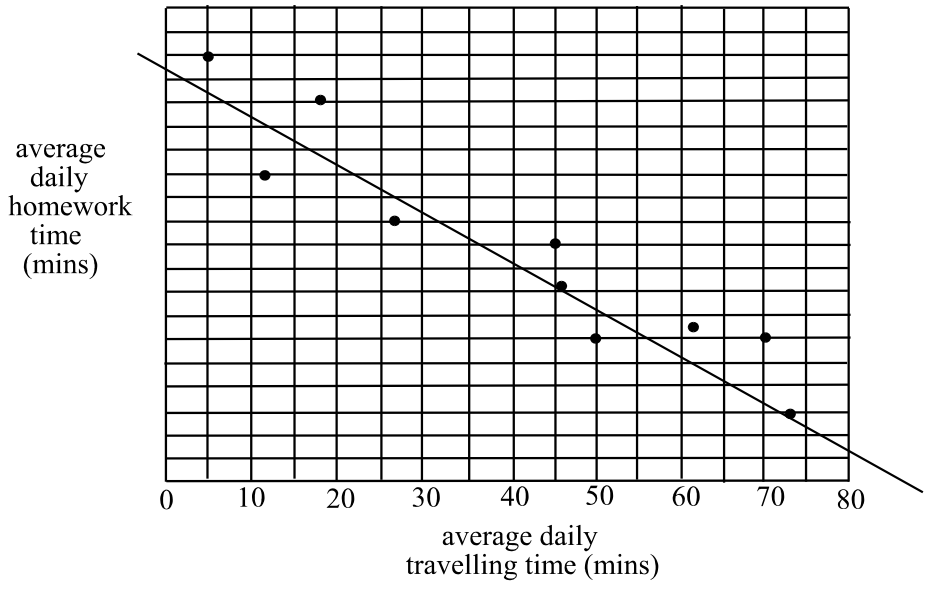


Diagram 1

b. Looking at the scatterplot in **Diagram 1**, it appears that as travelling time increases, homework time [ ] and therefore the sign of *r*; Pearson's correlation coefficient, would be [ ].

2 marks

c. The percentage of variation in the homework time that can be accounted for by the variation in travelling time is 89.27%.

Using this fact together with your response to part **b**. the value of *r* (Pearson's correlation coefficient) is [ ]. Express your answer correct to two decimal places.

1 mark

d. A least squares regression line of the form $y = a + bx$ is to be calculated for the data using the formulae

$$b = r\frac{s_y}{s_x} \text{ and } a = \bar{y} - b\bar{x}.$$

Given that $S_y = 47 \cdot 80$ and $\bar{y} = 97 \cdot 8$, and using your answers to parts **a.** and **c.** find the equation of the least squares regression line. Express each value correct to 2 decimal places.

average daily homework time = [ ] + [ ] × average daily travelling time

2 marks

Question 5

A sample of ten Year 10 students were similarly asked to estimate their average daily travelling time to and from school together with their average daily time spent doing homework.

A scatterplot of this data is shown in **Diagram 2** below.

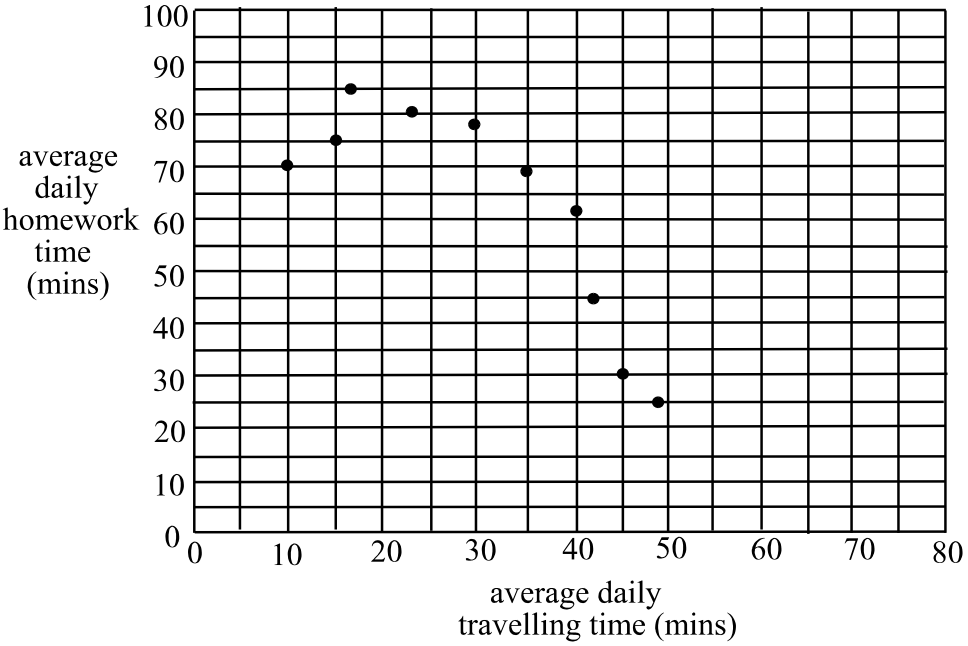


Diagram 2

A least squares equation was calculated for this data and was given by

$$\boldsymbol{average\ daily\ homework\ time = 100.42 - 1.26 \times average\ daily\ travelling\ time}$$

A residual plot for the data is shown in **Diagram 3** below.

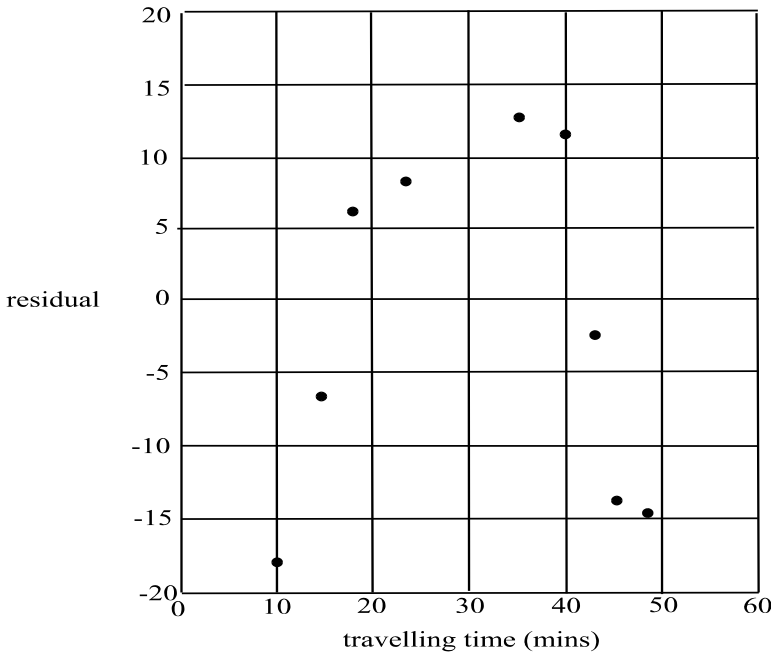


Diagram 3

a. Using the information in **Diagram 2** complete the residual plot in **Diagram 3** by filling in the missing point that relates to the student with a travelling time of thirty minutes.

1 mark

b. What aspect of the points plotted on the residual plot, in **Diagram 3**, suggest that a transformation is required to linearise the original data?

_____

_____

1 mark

A transformation where each of the values on the horizontal, (travelling time) axis is squared, is to be used in an attempt to linearise the data. The table is not complete and is shown as **Table 6** below.

Travelling time	10	15	17	23	30	35	40	42	45	48
$(\text{travelling time})^2$	100	225	289	529	900	1225	1600	1764	2025	
Homework time	70	75	85	80	78	69	62	45	30	25

Table 6

c. Complete **Table 6** above by filling in the entry in the shaded cell.

1 mark

d. In an attempt to linearise the data, what other transformation might reasonably have been used. Explain clearly which axis you would apply the transformation to.

1 mark

A scatterplot of average daily homework time (y) plotted against average daily travelling time squared, is shown in **Diagram 4** below.

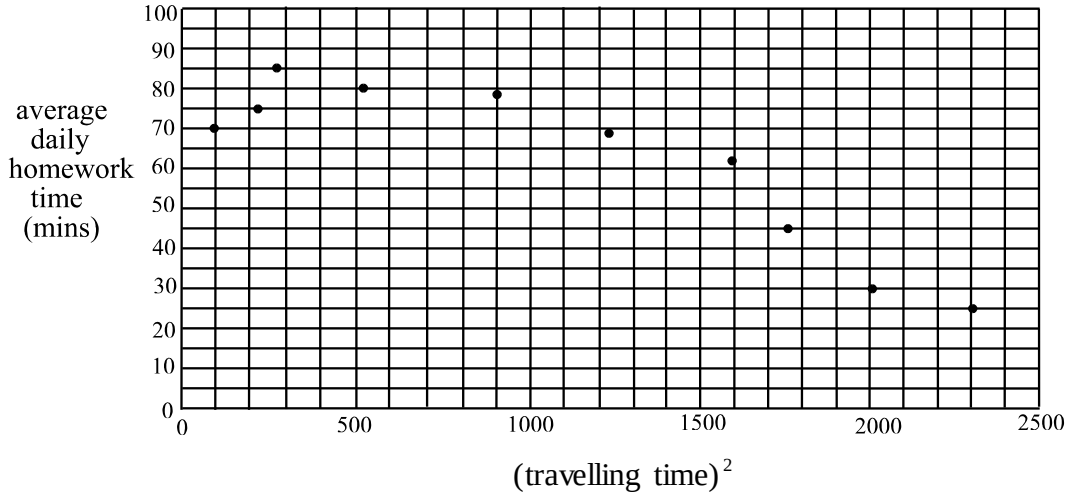


Diagram 4

e. Using **Table 6**, find the least squares regression line for average daily homework time against $(\text{travelling time})^2$. Express each value correct to two significant figures.

average daily homework time = ☐ + ☐ × (average daily travelling time)²

2 marks

f. To test whether this transformation has been successful in linearising the data

i. what should now be constructed?

ii. what feature should be present if the transformation has been successful?

1 + 1 = 2 marks