4. Kay is studying the variables Daily Total Sunshine (x) and Daily Total Rainfall (y) from the large data set for Leeming in 2015

Kay starts with 5th May and then selects every 10th day thereafter.

(a) State the name of the sampling technique Kay uses.

(1)

Kay wants to find the regression line of y on x for these data.

(b) Using your knowledge of the large data set, explain how Kay might need to clean these data before finding the equation of the regression line.

(1)

The equation of the regression line Kay finds is $y = 0.741 + 0.199x$

(c) Using your knowledge of the large data set,

(i) state the units of the gradient of the regression line,

(ii) give an interpretation of the y-intercept of the regression line.

(2)

Kay's teacher claimed that the greater the amount of sunshine in a day the lower the amount of rain there should be.

(d) State, giving a reason, whether the teacher's claim is true for Kay's data.

(1)

The teacher used all the data for these variables from the large data set for Leeming in 2015, as a sample.
The teacher calculated the product moment correlation coefficient for x and y to be -0.160

In a suitable test to determine whether there is evidence to support the teacher's claim, the p-value was 0.015

(e) Using a 5% level of significance, state the hypotheses and conclusion for this test.

(2)

(f) For the test in part (e) describe

(i) the sample,

(ii) a possible population.

(2)