

2. Marc took a random sample of 16 students from a school and for each student recorded

- the number of letters, x , in their last name
- the number of letters, y , in their first name

His results are shown in the scatter diagram on the next page.

(a) Describe the correlation between x and y . (1)
 (a) weak negative correlation
 (As x increases, y decreases, but the observations are not close to a straight line) (1 mark)

Marc suggests that parents with long last names tend to give their children shorter first names.

(b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer. (1)
 (b) Marc suggests negative correlation and the data shows negative correlation, so Marc's suggestion is compatible with the data. (1 mark)

The results from Marc's random sample of 16 observations are given in the table below.

x	3	6	8	7	5	3	11	3	4	5	4	9	7	10	6	6
y	7	7	4	4	6	8	5	5	8	4	7	4	5	5	6	3

(c) Use your calculator to find the product moment correlation coefficient between x and y for these data. (1)
 (c) There is evidence of negative correlation between length of student's first and last names. (1 mark)
 (give conclusion in context of problem)

(d) Test whether or not there is evidence of a negative correlation between the number of letters in the last name and the number of letters in the first name.

You should

- state your hypotheses clearly
- use a 5% level of significance

(d) For Product Moment Coefficient, table gives 0.4259
 Observed -0.5446 < -0.4259
 so we reject H_0 in favour of H_1 at 5% sig level (1 mark) (3)

(c) {fx-991EX: MENU6-Stats/ $y = a + bx$ / <Enter Data> / OPTN / Regression Calc }
 {fx-991EX: MENU2-Stats / <Enter Data in List 1 & List 2> / CALC / REG / }

$\Rightarrow r = -0.54458..$
 $= -0.545$ 3sf
 (1 mark)

SET to Check Settings. Should be
 2Var XList : List 1
 2Var YList : List 2
 then EXIT to return
 2Var Freq : 1
 X/a+bx

(d) $H_0: \rho = 0$ (null hypothesis always zero correlation for these)
 $H_1: \rho < 0$ (1 mark)

"Critical Values for Correlation Coefficients" Table in Formula Book gives one-tailed values, so 0.05 column is appropriate for 5% sig level.
 Sample size, $n = 16$, but values are for $p > 0$, so we are testing for <- value as signif.

