**The Correlation Formula and Regression**

1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | 3 | 5 | 8 | 4 |
| ***y*** | 2 | 6 | 9 | 3 |

The covariance from the data in the table above is sxy = 5.

(a) Use your GDC to calculate .

(b) Use your GDC to calculate .

(c) Calculate the product-moment correlation coefficient.

2. These are similar to Question 1. In each case you are given the value of sxy and need to find the correlation coefficient.

(a) sxy = –3.125

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | 8 | 4 | 7 | 2 |
| ***y*** | 4 | 9 | 6 | 7 |

(b) sPQ = 13.2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **P** | 2 | 4 | 6 | 8 | 10 |
| **Q** | 10 | 17 | 18 | 30 | 20 |

(c) = 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***x*** | 11 | 24 | 15 | 18 | 5 |
| ***y*** | 9 | 10 | 6 | 4 | 6 |

(d) = –6.889

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 15 | 7 | 20 | 11 | 4 | 8 |
| ***y*** | 2 | 5 | 1 | 4 | 3 | 7 |

3. (a) Plot the points from Question 2d on a scatter plot.

(b) Use your GDC to find the equation of the regression line for *y* on *x*.

(c) Plot this regression line on your scatter plot.

(d) Estimate to 1 decimal place the y-value if *x* = 17.

4. Use the table for Question 2c here. Find the values of,  and  from your GDC.

Without further use of your GDC, show that the equation of the regression line for *y* on *x* can be written as:

*y* = 0.0974*x* + 5.58.

5. The 100m and 400m times (in seconds) for seven athletes are given in the table:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **name** | A | B | C | D | E | F | G |
| **100m** | 11.7 | 12.8 | 13.0 | 11.2 | 13.4 | 13.6 | 12.1 |
| **400m** | 56.5 | 63.0 | 63.2 | 57.0 | 64.5 | 64.8 | 59.8 |

(a) Treating 100m as *x* and 400m as *y*, find the equation of the regression line for *y* on *x*.

(b) Athlete H has a time of 12.4s for the 100m. Use your answer to (a) to estimate his 400m time.

6. The percentage scores of 8 students in the Mathematics and Physics examinations are given below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **name** | A | B | C | D | E | F | G | H |
| **Math** (*x*) | 64 | 80 | 60 | 55 | 71 | 38 | 66 | 49 |
| **Phys** (*y*) | 50 | 77 | - | 51 | 65 | 49 | 62 | 50 |

Student C was absent for the Physics exam. Use the results of the other 7 students to find the equation of the regression line for *y* on *x*, then show that a reasonable estimate of C’s Physics score is 57 (correct to the nearest integer).

7. The length (*x* mm) and width (*y* mm) of a sample of leaves from a tree is given below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 70 | 55 | 87 | 46 | 94 | 51 | 49 | 66 | 80 | 52 |
| *y* | 26 | 20 | 32 | 20 | 31 | 19 | 15 | 26 | 25 | 15 |

(a) By obtaining the equation of the regression line for *y* on *x*, estimate the width of a leaf with length:

(i) 77mm (ii) 18mm.

Give answers correct to the nearest mm.

(b) Comment on the reliability of your answer to (a)(ii).