**TASK 1: QUANTIFYING CORRELATION**

**In-class investigation**

**Unit 3**

**Topic 3.1: Bivariate data analysis**

**Course-related information**

The concepts and skills developed in this investigation relate to the following dot points within the WA Mathematics Applications syllabus:

3.1.5 construct a scatterplot to identify patterns in the data suggesting the presence of an association

3.1.6 describe an association between two numerical variables in terms of direction (positive/negative), form (linear/non-linear) and strength (strong/moderate/weak)

3.1.9 use a scatterplot to identify the nature of the relationship between variables

**Background information**

For this investigation, it is not expected that students describe relationships as strong/weak, linear/non-linear or positive/negative; nor do they need to have studied correlation. However, they should be able to describe the relationships between variables in terms of “as variable one increases then the other increases/ decreases or does not show a pattern.” Previous experience with creating and interpreting scatterplots is assumed.

**Task conditions**

This task consists of an in-class investigation for which students might need 40 – 55 minutes to complete. Access to technology is not necessary for the investigation.

**Quantifying correlation - Bivariate data analysis**

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



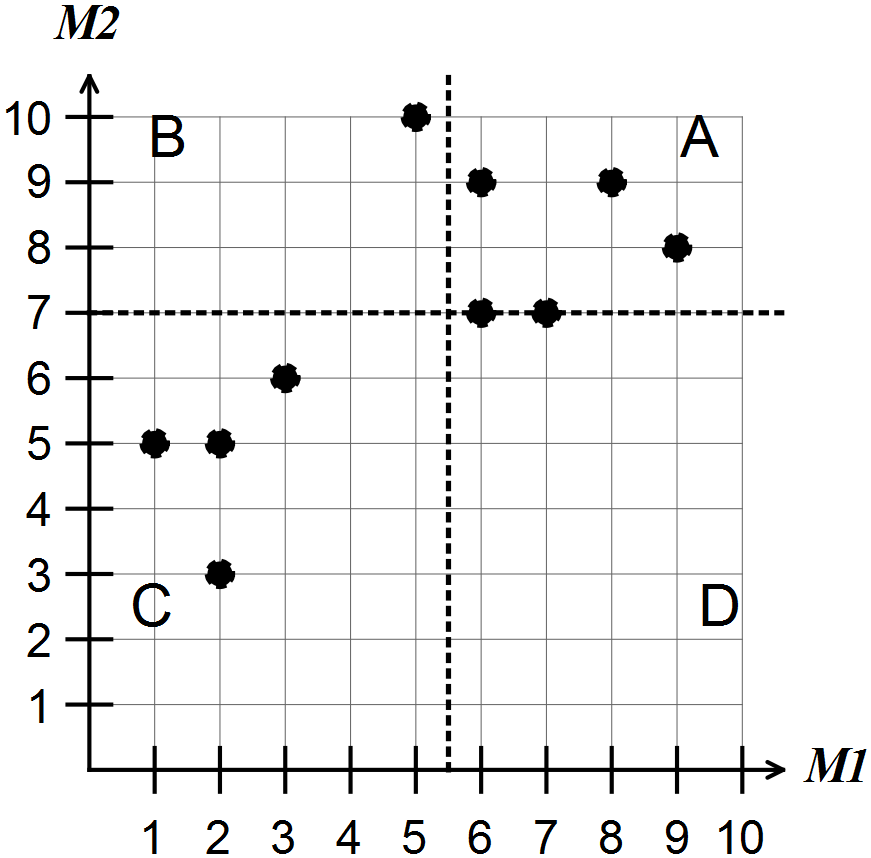
**In-class investigation 1 (Total marks: 50)**

In this investigation a process for measuring the relationship between two variables is examined. The measure is called the *q*-correlation coefficient.

To calculate the *q*-correlation coefficient

* Plot a scatter graph for the two variables.
* Determine the median for the values on the horizontal (*x*) axis.
* Determine the median for the values on the vertical (*y*) axis.
* Divide the region where the points occur into 4 quadrants by
  + Drawing a vertical line where *x* = median of values on the horizontal axis
  + Drawing a horizontal line where *y* = median of values on the vertical axis
* Label the quadrants e.g. A, B, C and D in the positions as shown on the graph below. The top right quadrant is always A, the top left is B, bottom left is C and bottom right is D.
* Count the points in each quadrant (do not count points on the dividing lines).
* Use the rule *q*-correlation coefficient = where *a* represents the number of points in quadrant A, *b* represents the number of points in quadrant B, *c* represents the number of points in quadrant C, and *d* represents the number of points in quadrant D.

Example: M1 and M2 represent the scores of ten students in two mental Maths tests.

****

For the scatter graph shown above

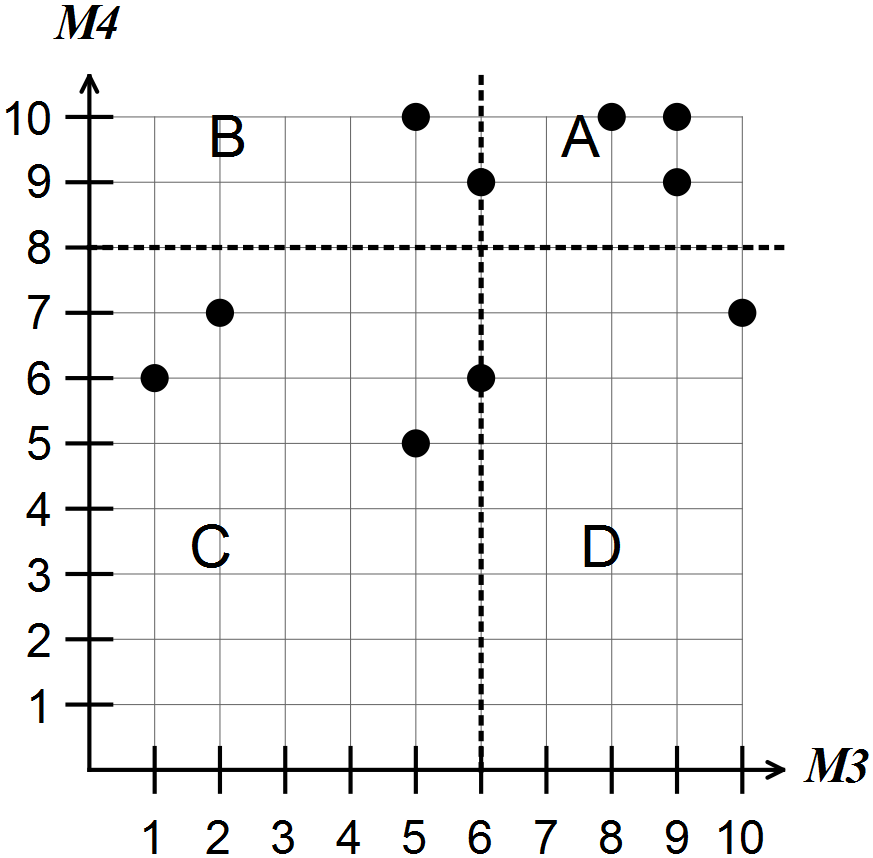
* the median of M1 is 5.5
* the median of M2 is 7
* there are 3 points in A so *a* = 3. Similarly *b* = 1, *c* = 4, *d* = 0.

The points (6,7) and (7,7) are on a line and not counted in any quadrants.

Hence, the *q*-correlation coefficient is equal to 

**Question 1 (7 marks)**

M3 and M4 represent students’ scores in Mental Maths tests

****

(a) How many students are represented? (1)

(b) What is the median for M3? (1)

(c) What is the median for M4? (1)

(d) How many students scored above the medians in both their tests?

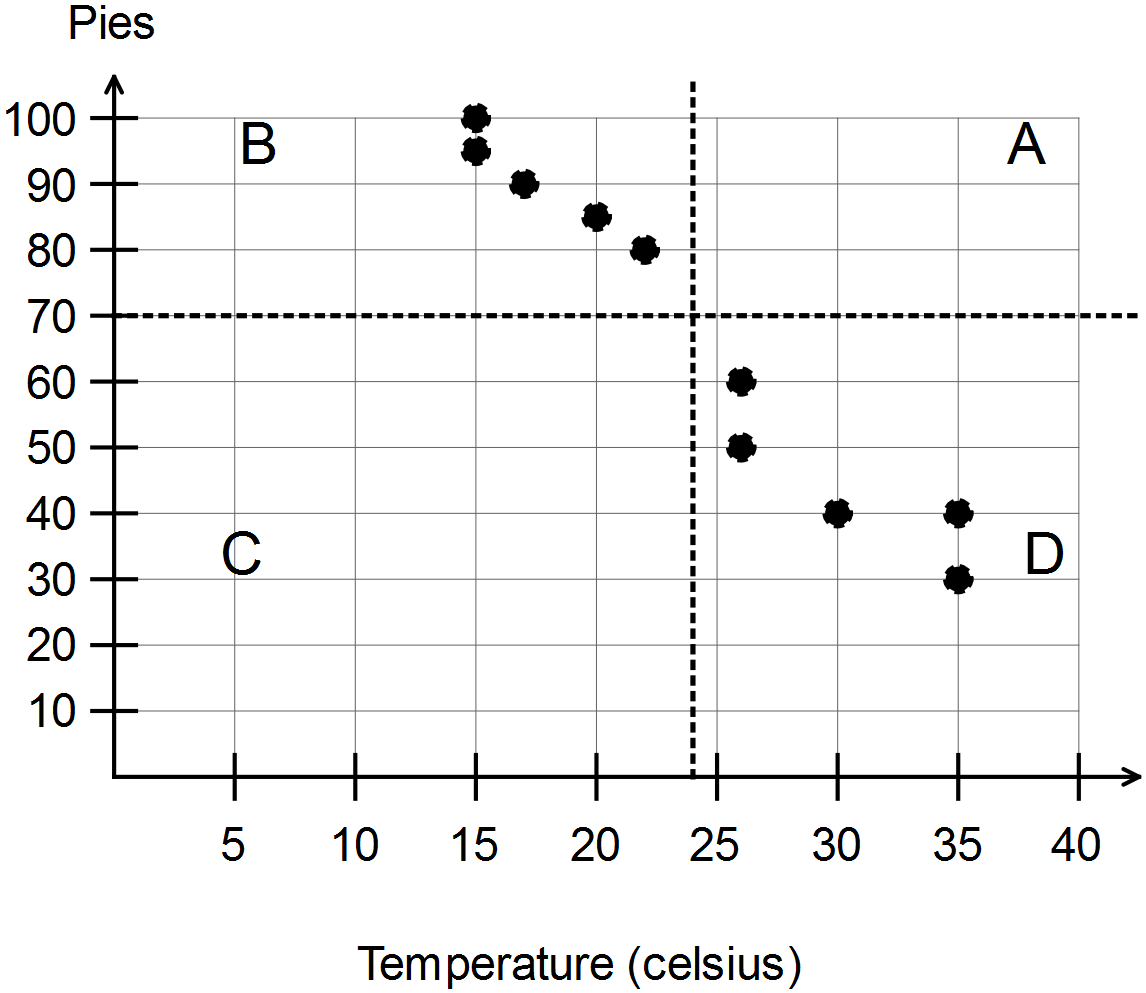
In what quadrant are the points for these students located? (2)

(e) Determine the*q*-correlation coefficient for the relationship between M3 and M4.

(2)

**Question 2 (7 marks)**

The graph shows a sample of daily maximum temperatures and the number of pies sold at the canteen during lunch time at one particular school.

****

(a) Describe the relationship between the number of pies sold and the daily maximum temperature. (1)

(b) Determine the median number of pies sold at the canteen. (1)

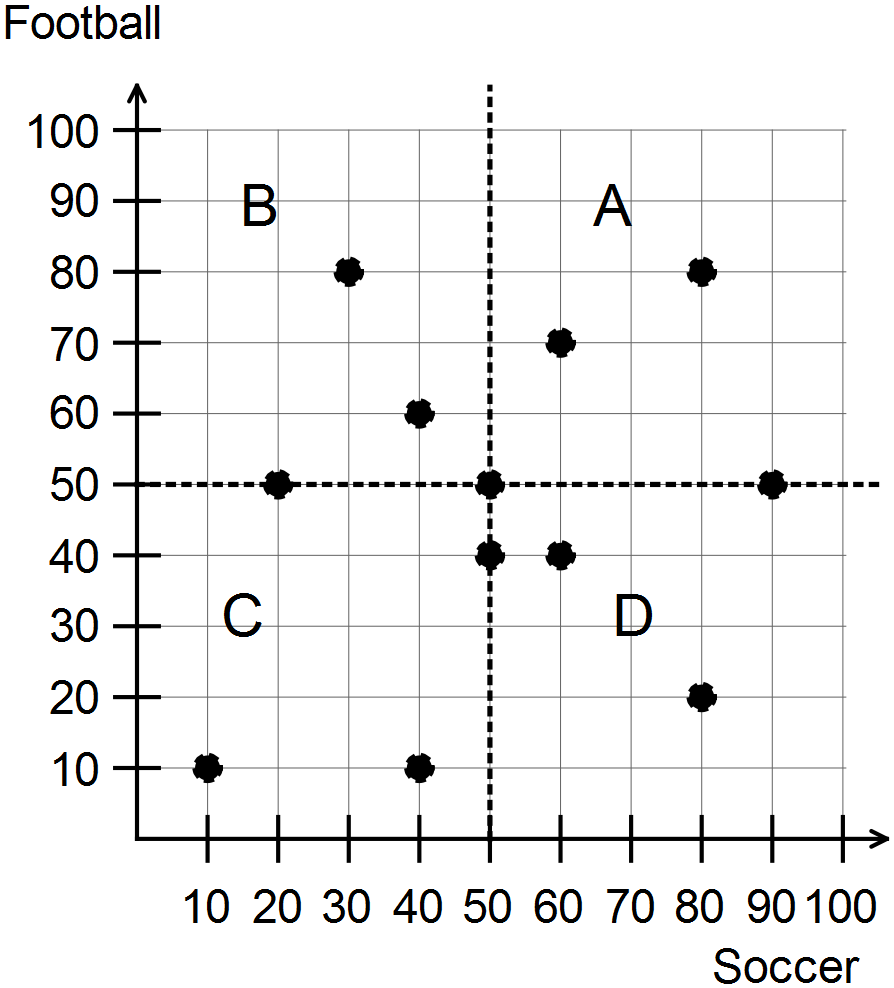
(c) Explain why there are no points on the dotted lines on this graph. (2)

(d) Determine the*q*-correlation coefficient for the relationship between the number of pies sold and the daily maximum temperature. (2)

(e) Determine the highest value that the *q*-correlation coefficient can take. (1)

**Question 3 (8 marks)**

Students in each of the Years 1 to 12 were asked if they liked two particular sports. The percentages of students in each class liking the two different sports are shown on the scatter graph below.

****

(a) What was the lowest percentage of students in any class liking football? (1)

(b) In any one class what was the highest percentage of students liking either sport?

(1)

(c) Describe what the point (50, 50) represents. (2)

(d) Determine the*q*-correlation coefficient for the relationship between the percentage of students liking football and the percentage of students liking soccer. (2)

(e) From the points on the scatter graph, what appears to be the relationship between the percentage of students liking football and the percentage of students liking soccer in each class? Justify your answer. (2)

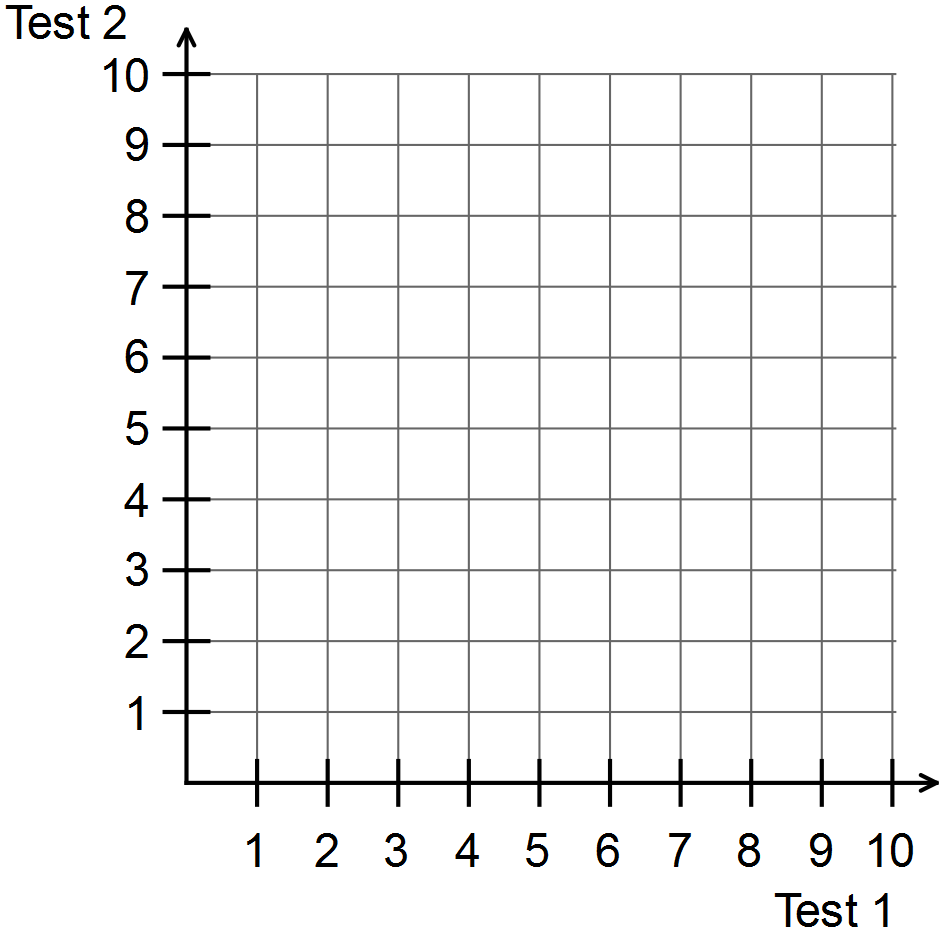
**Question 4 (7 marks)**

In this question the process of determining the four quadrants will be further examined.

The following table shows the mental maths scores (out of 10) of two different tests for 10 students.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Student** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| **Test 1** | 1 | 2 | 6 | 6 | 7 | 10 | 8 | 5 | 5 | 8 |
| **Test 2** | 2 | 3 | 5 | 4 | 5 | 9 | 7 | 6 | 4 | 8 |

(a) Represent these data as a scatter graph on the axes below. (2)

****

(b) Calculate the median of the Test 1 scores. (1)

(c) Calculate the median of the Test 2 scores. (1)

(d) Draw a vertical line to represent the location of the median for Test 1.

[Hint: consider the vertical lines in questions 1-3] (1)

(e) Draw a horizontal line to represent the location of the median for Test 2.

[Hint: consider the horizontal lines in questions 1-3] (1)

(f) Label the four quadrants as A, B, C and D. (1)

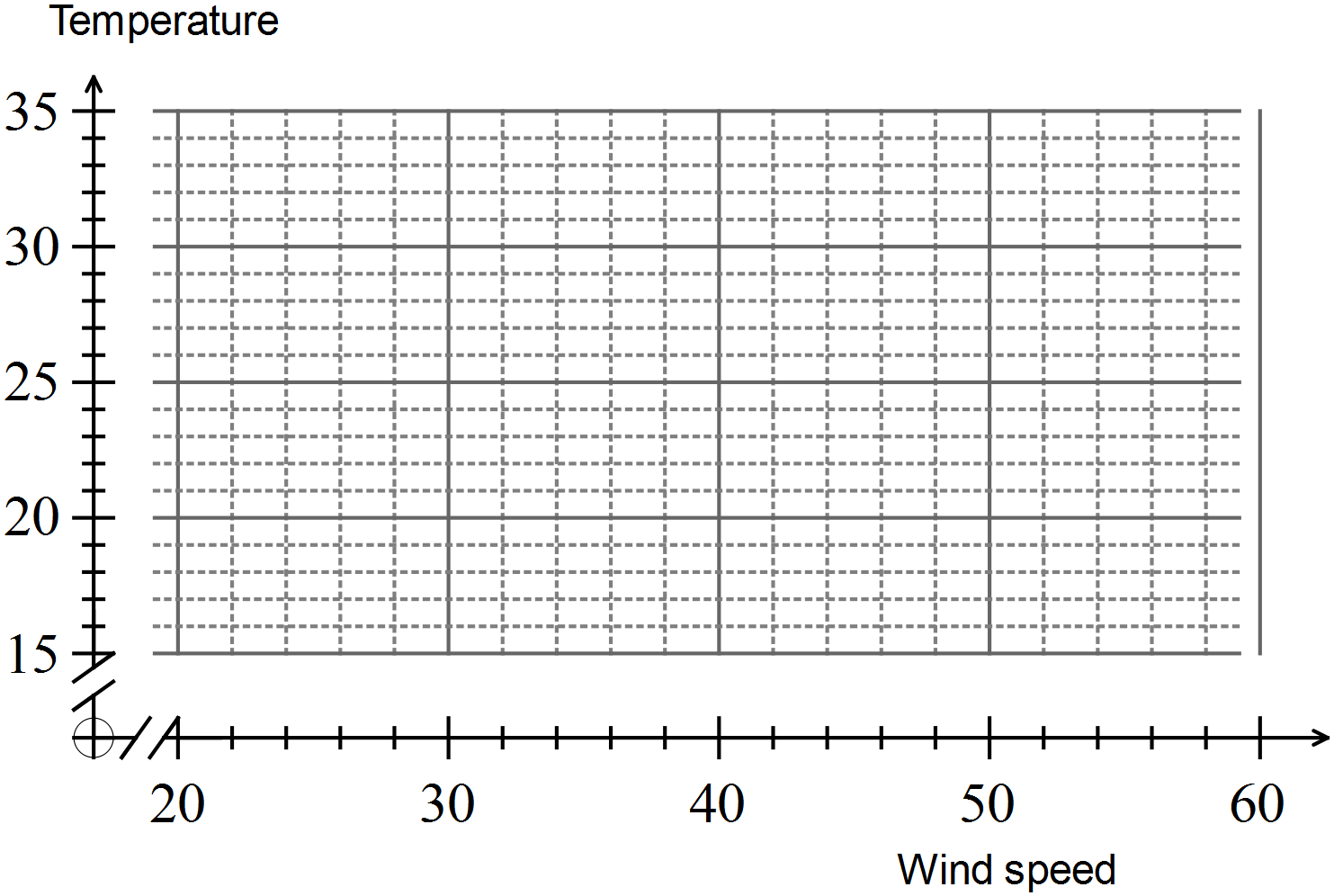
**Question 5 (12 marks)**

The following table shows the maximum wind speeds (km/h) and the maximum temperatures (degrees Celsius) for the first day of each of the last 12 months.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Month** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **Wind speed** | 31 | 22 | 22 | 33 | 30 | 31 | 33 | 35 | 39 | 57 | 43 | 28 |
| **Maximum**  **temperature** | 22 | 21 | 23 | 24 | 24 | 29 | 31 | 25 | 29 | 34 | 31 | 28 |

Use the process described in the previous questions to

(a) draw a scatter graph to show the relationship between the wind speed and the maximum temperature (3)



(b) Describe the relationship between maximum wind speed and maximum temperature referring to the location of the points to justify conclusions.

(3)

(c) divide the region of the plotted points into 4 labelled quadrants (4)

(d) calculate the *q*-correlation coefficient for the relationship between maximum wind speed and maximum temperature. (2)

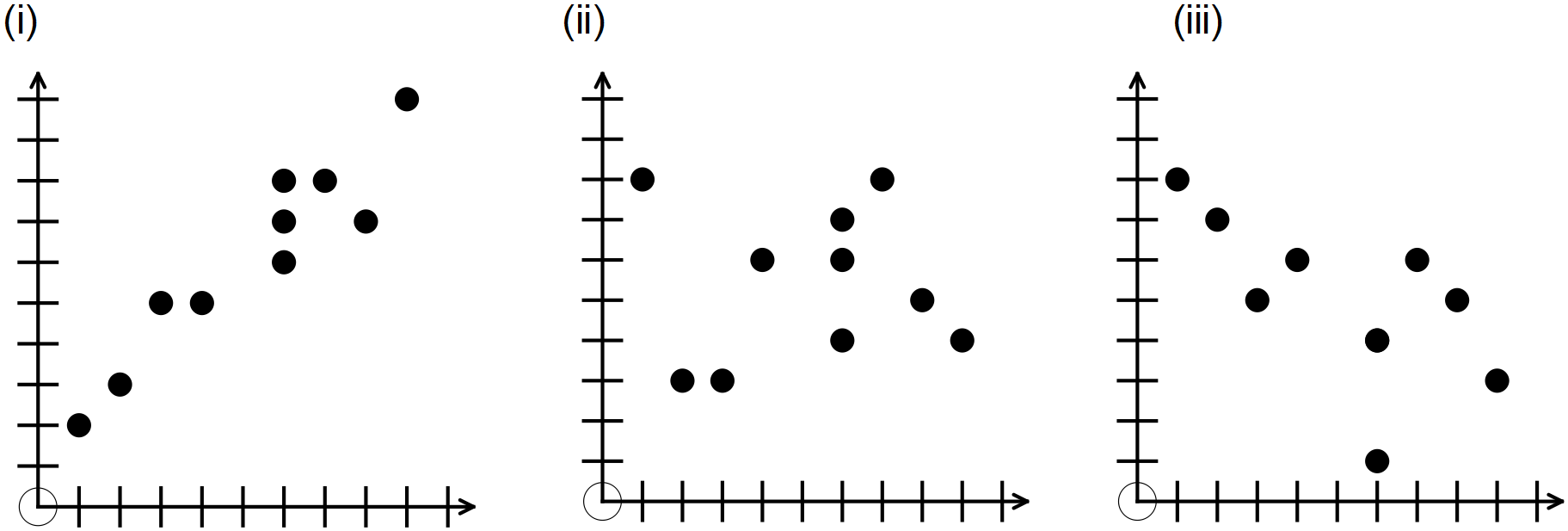
**Question 6 (9 marks)**

(a) You are given a set of possible values for the *q*-correlation coefficient.

(6)

-1 -0.8 -0.6 0.1 0.5 0.9

Which of these values is most likely in each of the six situations given?



(iv) there is little or no relationship between the two variables

(v) there is a very strong positive linear relationship

(vi) as one variable increases the other decreases and the points lie in a straight line.

(b) State one possible advantage of using the *q*-correlation coefficient to determine the strength of a relationship between two variables. (1)

(c) State two possible disadvantages of using the *q*-correlation coefficient to determine the strength of a relationship between two variables. (2)

**End of questions**

**Quantifying correlation**

**In-class investigation**

**Solutions and marking key**

**Question 1**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Solution | Mathematical behaviours | Marks |
| (a) | 10 | * reads number of points on scatter graph | 1 |
| (b) | 6 | * reads position of horizontal median line | 1 |
| (c) | 8 | * reads position of vertical median line | 1 |
| (d) | 3 in quadrant A | * interprets values on scatter graph * identifies location of points | 1  1 |
| (e) | *q*-correlation coefficient  = (3 + 3 - 2) ÷8  = 0.5 | * reads number of points on scatter graph * calculates *q*-correlation coefficient | 1  1 |

**Question 2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Solution | Mathematical behaviours | Marks |
| (a) | As the temperature rises the number of pies sold decreases | * interprets scatter graph | 1 |
| (b) | 70 | * reads position of horizontal median line | 1 |
| (c) | There are no days where the temperature was 24 nor where the number of pies sold was 70 | * identifies median not a score for temperature * identifies median not a score for pies sold | 1  1 |
| (d) | *q*-correlation coefficient  = )(0 + 0) – (5 + 5)) ÷ 10  = -1 | * reads number of points on scatter graph * calculates *q*-correlation coefficient | 1  1 |
| (e) | 1 | * interprets calculation of *q*-correlation coefficient | 1 |

**Question 3**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Solution | Mathematical behaviours | Marks |
| (a) | 10% | * reads scatter graph | 1 |
| (b) | 90% | * reads scatter graph | 1 |
| (c) | The class in which 50% liked soccer and 50% liked football. | * interprets scatter graph * describes both variables | 1  1 |
| (d) | *q*-correlation coefficient  = ((2 + 2 – (2+2)) ÷8 = 0 | * reads number of points on graph * calculates *q*-correlation coefficient | 1  1 |
| (e) | No relationship. The points are scattered randomly. | * Describes lack of relationship * interprets scatter graph | 1  1 |

**Question 4(a)(d)(e)(f)**

|  |  |
| --- | --- |
| Solution | |
|  | |
| Mathematical behaviours | Marks |
| * plots most points accurately * plots all points accurately * draws vertical line at median for Test 1 * draws horizontal line for median at Test 2 * labels quadrants as previously instructed | 1  1  1  1  1 |

**Question 4(b)**

|  |  |
| --- | --- |
| Solution | |
| Median = 6 | |
| Mathematical behaviours | Marks |
| * calculates median | 1 |

**Question 4(c)**

|  |  |
| --- | --- |
| Solution | |
| Median = 5 | |
| Mathematical behaviours | Marks |
| * calculates median | 1 |

**Question 5(a)**

|  |  |
| --- | --- |
| Solution | |
|  | |
| Mathematical behaviours | Marks |
| * plots coordinates for each point in correct order * plots most points accurately * plots all points accurately | 1  1  1 |

**Question 5(b)**

|  |  |
| --- | --- |
| Solution | |
| As wind speed increases so does the temperature; the points are generally rising to the right | |
| Mathematical behaviours | Marks |
| * notes increasing wind speed is associated with increasing temperature * describes location of points | 2  1 |

**Question 5(c)**

|  |  |
| --- | --- |
| Solution | |
|  | |
| Mathematical behaviours | Marks |
| * determines median maximum wind speed * determines median maximum temperature * draws median lines and labels quadrants | 1  1  2 |

**Question 5(d)**

|  |  |
| --- | --- |
| Solution | |
| *q*-correlation coefficient = (8 - 4) ÷ 12 = | |
| Mathematical behaviours | Marks |
| * determines number of points in each quadrant * calculates the *q*-correlation coefficient | 1  1 |

**Question 6**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Solution | Mathematical behaviours | Marks |
| (a) | (i) 0.9  (ii) 0.1  (iii) -0.6  (iv) 0.1  (v) 0.9  (vi) -1 | * identifies positive correlation from graph * identifies lack of relationship from graph * identifies negative correlation from graph * identifies lack of correlation from description * identifies positive correlation from description * identifies negative correlation from description | 1  1  1  1  1  1 |
| (b) | Easy to follow/calculate  More precise than guessing | * suggests one advantage | 1 |
| (c) | Inaccurate  Time-consuming | * suggests disadvantage of the process * suggests disadvantage of the process | 1  1 |