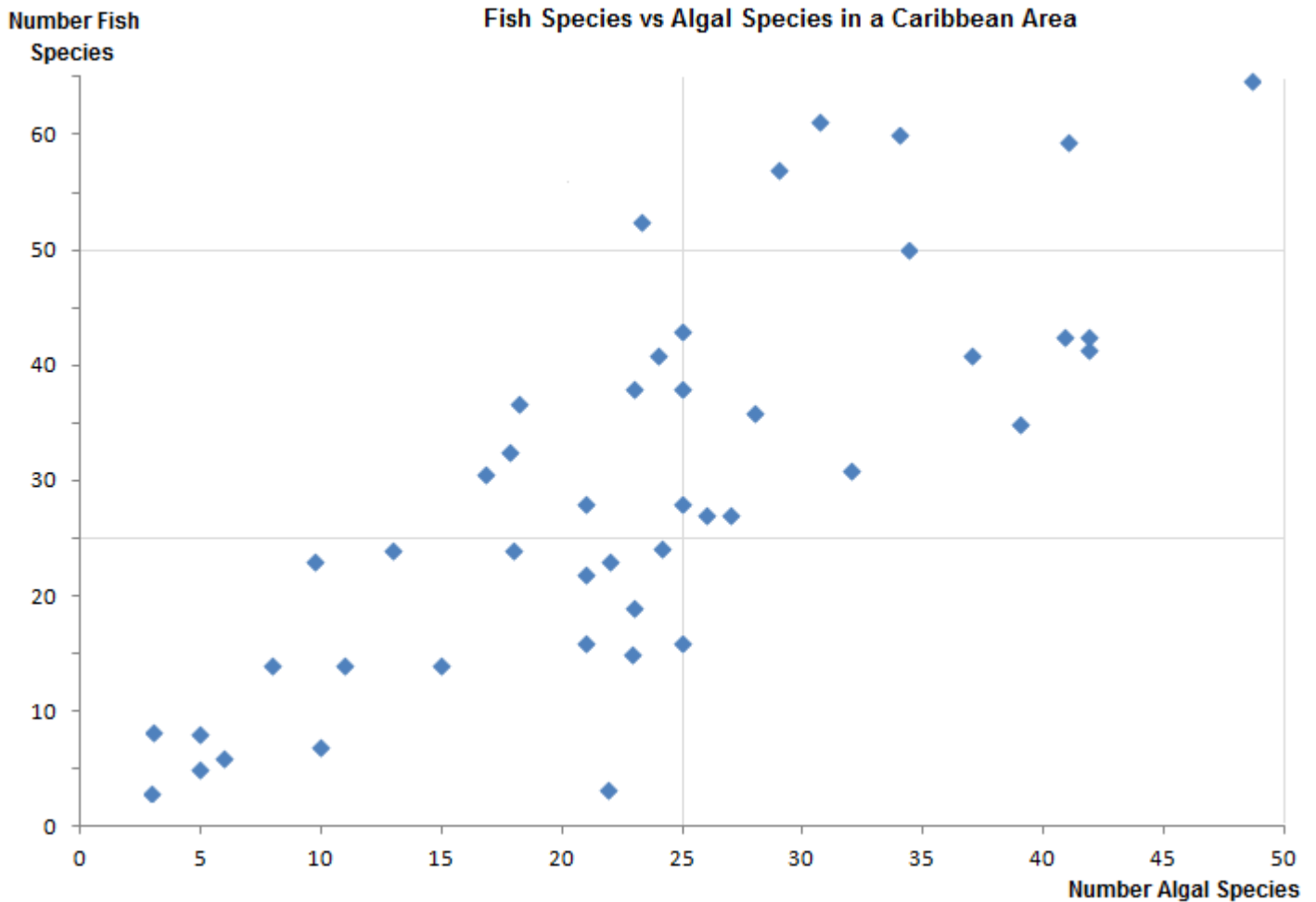## Level 1 Data Practice #5

Biologists in a Caribbean country are interested in what affect various factors have on the variety of local fish.
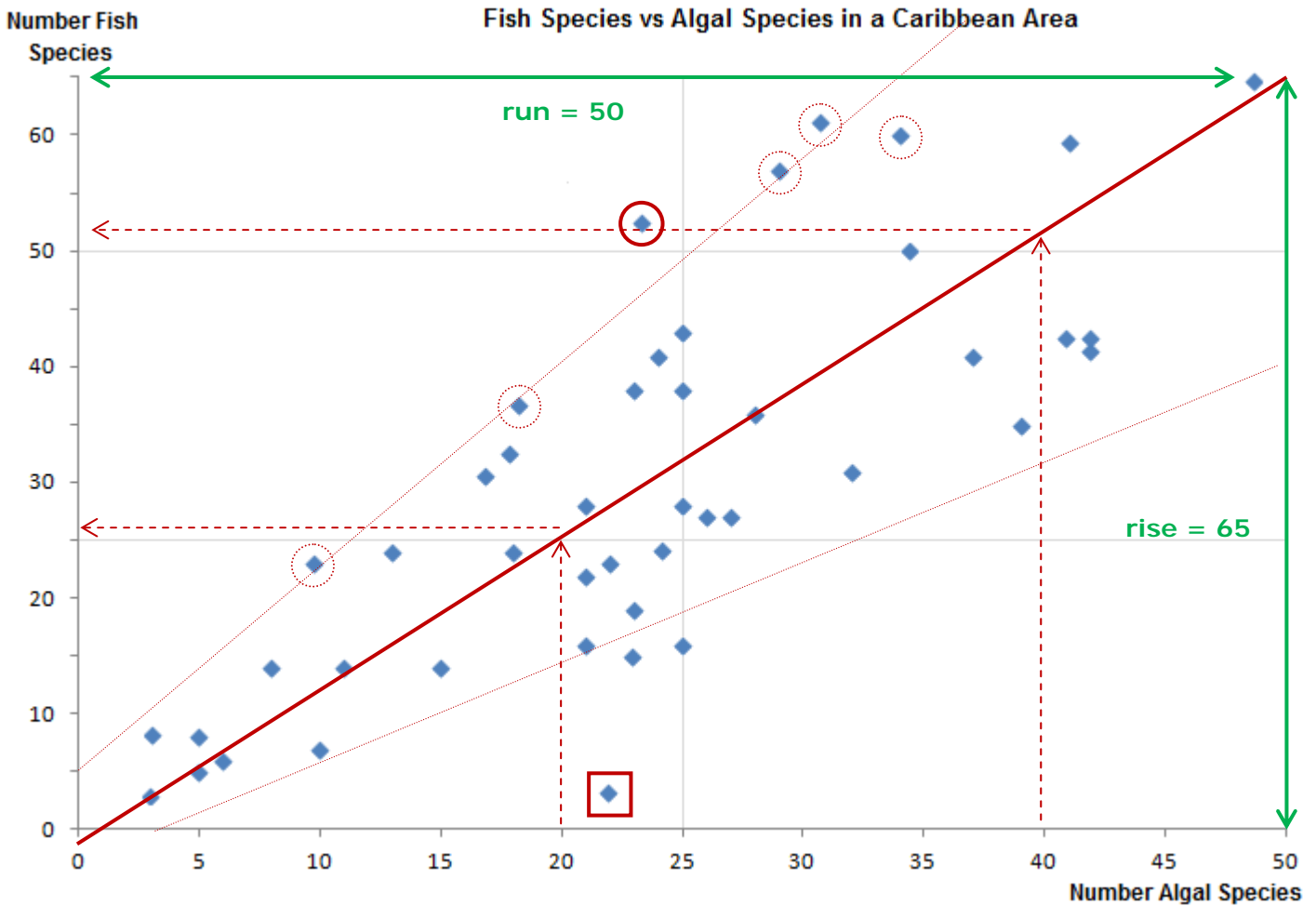
They conduct a survey that includes measuring the number of species of algae in an area and the number of different species of fish seen in that area.

The results of that part of the study are shown below.

Fish Species vs Algal Species in a Caribbean Area

1. Describe the relationship between the number of fish species seen and the number of algae species seen.

2. Circle a point (or points) on the graph which have more fish than would be predicted from the amount of algae present.

   Put a square around any point which might be considered an outlier to the relationship.

3. Predict the amount of fish species you would predict to see if you were in an area with 20 algae species, and then for an area with 40 algae species present.

   How reliable do you believe your predictions are?

4. A scientist reporting on the data says "The presence of more algae in an area causes there to be more fish". Does the data support this?

5. Construct a model for the number of fish species seen from the amount of algae species seen.

Answers: Level 1 Data Practice #5

Fish Species vs Algal Species in a Caribbean Area

Number Fish Species (vertical axis), *Number Algal Species* (horizontal axis)

run = 50, rise = 65

1. As the number of algal species seen increases there are more fish species seen, although there is quite a lot of variation from that general pattern. Generally just over one extra fish species is seen for every algal species seen. Technically – there is a fairly weak positive linear correlation between them.

   (Try to put numbers to the relationship. Note there is some scope for differences with strength, in this case "weak" and "moderate" etc are acceptable. But not "strong".)

2. The bold circle point must be included, plus any of the lightly circled points, depending on how the difference from the pattern is judged. Note these are not the warmest or coldest, but the points furthest above and below the general pattern shown by the line.

   The squared point is the only outlier – it is easily the furthest from the line of best fit.

3. 20 algal species gives a prediction of 26 (approx) fish species. 40 algal species gives a prediction of 52 (approx) fish species. You should show the line of best fit for this.

   The predictions are likely to be poor, as the correlation is weak. The one at 40 algae is even less certain than the one at 20, as the relationship gets weaker with more species.

4. This statement cannot be supported. Firstly a scatter plot relationship **never** says anything firm about what causes what, at best it might be evidence for it. Secondly the measures used are the number of species, not the amount of actual algae or fish.

5. $F = \frac{65}{50} A - 1$ or close mathematical equivalents, depending on the line of best fit drawn.

   where F = number fish species seen, and A = number algae species seen