

## Answer Key - Getting to Know the Regression Line

1. There are infinitely many lines through any one point in the plane. The computer does not know which one to draw.
2. In addition to errors in calculation and algebra, there may be differences because of putting the point in slightly the wrong place on the graph. The points are fat and cover a lot of territory. Make sure when clicking or dragging a point that the small circle around the point has appeared.
3. There would be one line that contains all three points if the student had accidentally drawn them in a straight line to begin with. Unless students did this on purpose, it is unlikely, so there will be no line that contains all of them. A line is determined by two points. The third point may or may not be on that line.

Answers may vary for the comparison of their line with the computer's line. Note that the line fitting the points does not need to pass through any of the points. It just needs to show the overall trend.

4. Answers will vary. When a new point is added to the scatterplot, the line will tend to move over closer to that new point. Added points that are farther away from the previous line will make it move further than added points that are closer. Adding a point that is very far away from all the other points can have a dramatic effect. There will be no effect on the line only if the new point is already on the line. If there are already many, many points in the plot, the new point may not have much effect at all.

Adding several points at a time will have no effect if the new points are scattered in a balanced way around the old line, even if they are far away. But most of the time the new points will be unbalanced around the old line, so the line will move in the direction of the new points that are farther away if there are many of them.

5. The answers will vary, but you will see whether they are correct by whether the line is horizontal or not.

**a)** There can be many, many answers. One example is determining the "ideal" weight for your height by fitting a line to a scatterplot of height and weight measurements for a large sample of people in the general population. Another example could be predicting the grade you might get on a test as a function of how many hours you spent studying for the test. Or predicting the moisture content of a synthetic fibre in a storehouse as a function of relative humidity.

**b)** The error is the vertical distance between the point and the line: the distance between the data point and a point on the line directly below or above that data point. The computer squares these distances for each data point, adds all the squares together, and then takes the line for which this sum of squares is minimum. The computer doesn't actually calculate the sum of squares for all possible lines and then take the smallest. That would be impossible. It uses formulas for estimating the slope and y-intercept of the line of best fit from the data points. These formulas are derived using calculus.