Let's consider the 40 pairs of shoe print lengths and heights included in Data Set 2 of Appendix B. Those 40 pairs of values result in this regression equation: $\hat{y} = 80.9 + 3.22x$. The slope of 3.22 tells us that if we increase x (the length of the shoe print) by 1 cm, the predicted height of the person will increase by 3.22 cm. That is, for every additional 1 cm increase in the length of a shoe print, we expect the height to increase by 3.22 cm.

Outliers and Influential Points

A correlation/regression analysis of bivariate (paired) data should include an investigation of *outliers* and *influential points,* defined as follows.

DEFINITIONS

In a scatterplot, an **outlier** is a point lying far away from the other data points.

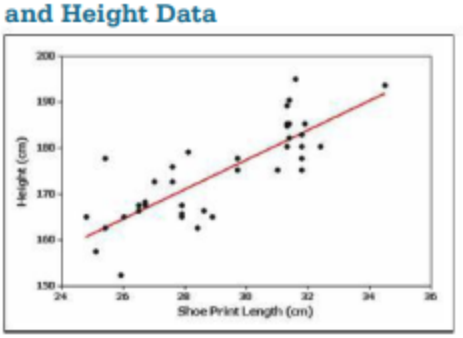
Paired sample data may include one or more **influential points,** which are points that strongly affect the graph of the regression line.

To determine whether a point is an outlier, examine the scatterplot to see if the point is far away from the others. Here's how to determine whether a point is an influential point: First graph the regression line resulting from the data with the point included, then graph the regression line resulting from the data with the point excluded. If the regression line changes by a considerable amount, the point is influential.

Example 5 Influential Point

Consider the 40 pairs of shoe print lengths and heights from Data Set 2 in Appendix B. The scatterplot located to the left below shows the regression line. If we include this additional pair of data: $x = 35$ cm, $y = 25$ cm (shoe length is 35 cm and height is 25 cm), we get the regression line shown to the right below. The additional point (35 cm, 25 cm) is an influential point because the graph of the regression line did change considerably, as shown by the regression line located to the right below. Compare the two graphs and you will see clearly that the addition of that one pair of values has a very dramatic effect on the regression line, so that additional point is an influential point. The additional point is also an outlier because it is far from the other points.

Original Paired Shoe Print and Height Data



Shoe Print and Height Data with an Additional Point (35 cm, 25 cm)