

STAT 213: INFLUENTIAL POINTS IN REGRESSION

The goal of this activity is to try to discover through experimentation what properties of a data point make it more or less influential on the estimates of the parameters of a regression line.

To get started, visit <http://lock5stat.com/statkey>, and choose **Descriptive Statistics for Two Quantitative Variables** from the left-hand side. From the drop-down menu in the upper left, select **Election Margin-2e (Margin vs. Approval)**, which contains data about presidential first term approval ratings and their subsequent reelection margins. Check the box in the lower right to show the best fitting regression line.

- (1) Explore the impact that an individual response value has on the intercept and slope of the regression line by clicking and dragging individual points *up* and *down*, while trying to hold the horizontal position constant (in essence, we are imagining that a particular election had gone differently, and asking how that would have affected our estimate of the relationship between approval and reelection margin). Try this for a few different points. Do different points have a different impact on the slope per unit movement? What seems to impact this rate?

- (2) Now explore the impact that the presence or absence of a particular data point has on the parameter estimates. Unfortunately the interface does not support selecting and deleting data points, but you can do this by clicking “Edit Data” and deleting a row from the data table. Take note of how much the slope changes when you delete each point, and reload the full data before you delete the next one. What characterizes the data points that have the biggest impact on the slope, in particular?