

Lab 5: Java applets for understanding regression
STT 422: Summer, 2004
Vince Melfi

There are some nice java applets related to understanding regression and correlation available on the web. Today we'll use two of these to improve our intuition about least squares and influential points. You'll need to start *Netscape Navigator* or *Internet Explorer* before beginning.

Least squares by eye

It's easy to think that the least squares line will be close to the best-fitting line chosen "by eye." The first java applet should disabuse you of this notion.

1. Go to http://www.ruf.rice.edu/~lane/stat_sim/reg_by_eye/index.html, and read the instructions.
2. Click on the "Begin" button. This should open a new window with a scatter plot of points.
3. We'll start by getting comfortable with the applet. By clicking the mouse and dragging, draw a line on the plot. The MSE will appear in the small box to the right of the scatter plot. The MSE is the average squared deviation: for the line $y = a + bx$,

$$\text{MSE} = \frac{1}{n} \sum (y_i - [a + bx_i])^2.$$

So the line that minimizes the MSE is the least squares line. Draw a few more lines on the plot, some that you think fit the data well and some that you think don't fit the data well, and look at the associated MSEs. Now click the box next to "Draw regression line" to see the least squares line. Also click "Show Minimum MSE" to see the MSE of the least squares line. Next guess the value of the correlation r by choosing one of the 5 values given, and then click "Show r " to see whether you got it right.

4. Now click on "New Data" to get a new scatterplot. (Sometimes this garbles the values for r ; if this happens, just minimize then maximize the window to get them back.) Draw what you think is the least squares line on the plot, and write the associated MSE in Table 1 below. Then have the applet draw the real least squares line and show the minimum MSE, and write the minimum MSE in Table 1. Also try to guess r and write your guess and the true value in Table 1. Repeat this (get new data, draw line, etc.) enough times to fill in Table 1. Do you still think it's easy to find the least squares line by eye?

Your MSE	Minimum MSE	Your r	True r

Table 1: Least squares by eye

Influential points

The next applet allows us to see the effect of an added point on the least squares line and correlation.

1. Go to <http://www.stat.uiuc.edu/~stat100/java/guess/PPApplet.html>. Read the description at the top of the page.
2. Change the “Target correlation” at the bottom of the plot to 0.9. Then click on “Random points” which should add 10 points with approximate correlation 0.9 to the plot. Click next to “Regression line” to get the least squares line plotted. Write the correlation and the equation of the least squares line in Table 2 in the row labeled “Original.”
3. Now add a point (by clicking with the mouse) that is at about the center of the points horizontally (so it’s reasonably near to \bar{x}) but is far from the points vertically. Write the new correlation and least squares line in Table 2 in the row labeled “Near \bar{x} .”
4. Erase the point you added. Do this by clicking next to “Erase point” and then clicking on the point you added. You should now have the original data with the original correlation and regression line. Now add a point that is near the right edge of the plot (so it is far from \bar{x}) and is about as far from the data vertically as the point you added previously. Again write the new correlation and least squares line in Table 2, this time in the row labeled “Far from \bar{x} .”
5. Which point changed the correlation and least squares line more?

	Correlation	Least Squares Line
Original		
Near \bar{x}		
Far from \bar{x}		

Table 2: Influence of added points when $n = 10$

- Click on “Clear plot” to remove all the points. Change the number of points from 10 to 100, and click on “Random points” to get a plot of $n = 100$ points with approximate correlation $r = 0.9$. Repeat the above steps of adding points, etc., writing the results in Table 3.
- What difference do you notice when $n = 100$ rather than $n = 10$?

	Correlation	Least Squares Line
Original		
Near \bar{x}		
Far from \bar{x}		

Table 3: Influence of added points when $n = 100$