Week 9: Discussion

David Diez

March 4, 2008

Regression assumptions ::

- Linearity — the fit must be linear (non-linear fits are beyond this course's scope)
- Independence — each observation must be independent of the others
- Equal variance & normality — it is assumed that the residuals are from a normal distribution with constant variance

Sample scatterplots :: how some issues will look:

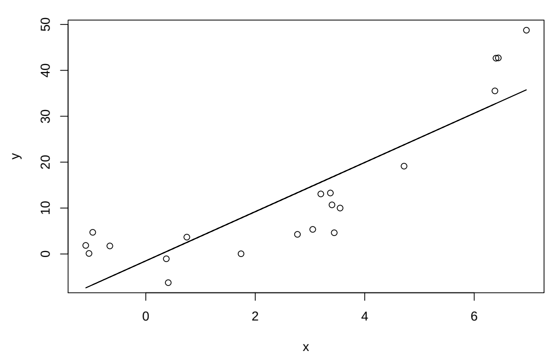

Figure 1: not a straight line – there is some curvature.

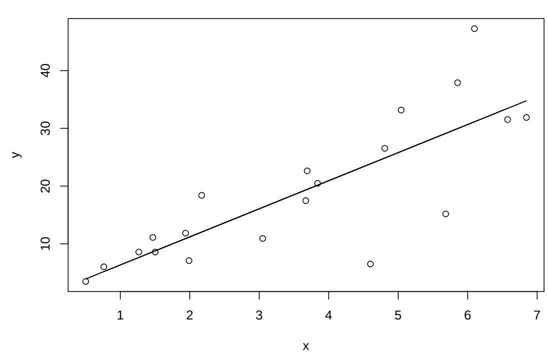

Figure 2: non-constant variance.

Properties of a LSR fit ::

- the sum of the residuals* equals zero
- take the mean of the x-values, $\bar{x}$, and the mean of the y-values, $\bar{y}$, then $(\bar{x}, \bar{y})$ will fall on the line (very important property!!!)
- the sum of the squared residuals is minimized by this line

* the residuals are the differences of the actual y-values from their regression line (ie, take each point and subtract off the y-value on the line corresponding to that value):

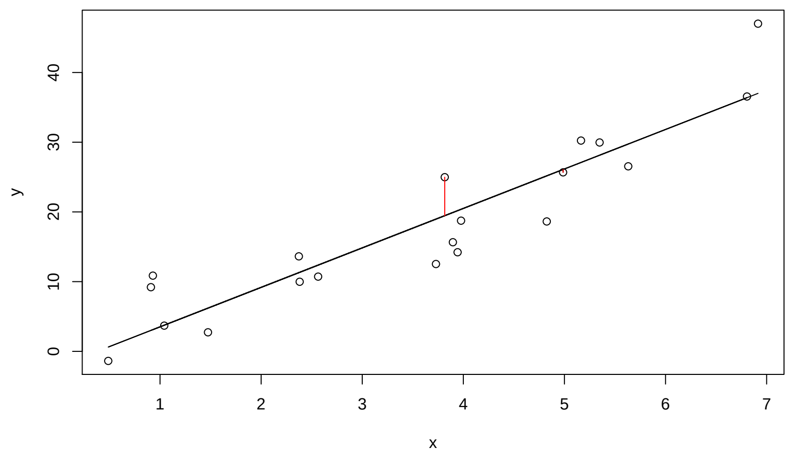

Figure 3: the vertical line above represents how to think of a residual – this residual is the length of this line and, since the observation falls above the line, is positive.

Review quiz #6, attempt 1 :: a useful equation:

$$slope = r\frac{s_y}{s_x}$$

Finding the LSR line :: assuming $\bar{x}, \bar{y}, s_x, s_y$, and r are provided,

1. Find the slope: $b_1 = r\frac{s_y}{s_x}$.
2. Use the fact that $(\bar{x}, \bar{y})$ is on the line to find the equation:

$$y - \bar{y} = b_1(x - \bar{x})$$
$$y = b_1 x - b_1\bar{x} + \bar{y} = [\bar{y} - b_1\bar{x}] + b_1 x$$

Note: the y-intercept is $\bar{y} - b_1\bar{x}$.