1. Don’t fit a line if a line doesn’t fit!

2. Always plot the data. Examples.

3. Better yet: always plot the residuals: \( e_i = y_i - \hat{y}_i \).

4. Plot the standardized residuals: \( e^*_i = e_i / s_{e_i} \). The standard error of the residual is given by \( s_{e_i} = s \sqrt{1 - h_i} \). The \( h_i \) are called the hat values and are given by:

\[
    h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}.
\]

5. Leverage: points far away from \( \bar{x} \) pull the regression line toward them. A way to measure leverage is by \( h_i \).

6. Influence: the amount that a point affects the regression line.

7. One way to measure influence is to refit the line with the point removed and see how things change. (R functions \texttt{dfbetas()}, \texttt{dffits}().)

```r
> l=lm(Gross~Budget,data=mov)
> summary(l)

Call:
  lm(formula = Gross ~ Budget, data = mov)

Residuals:
     Min      1Q  Median      3Q     Max
-121.72  -39.35   -9.55   28.18  203.22

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.622     19.665   1.05   0.300
Budget       1.382      0.216   6.41   2.9e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1

Residual standard error: 67.8 on 33 degrees of freedom
Multiple R-squared: 0.554,    Adjusted R-squared: 0.541
F-statistic: 41 on 1 and 33 DF,  p-value: 2.94e-07
```

> plot(residuals(l)~fitted(l))
> plot(rstandard(l)~fitted(l))
> hatvalues(l)
> dfbetas(l)
> dffits(l)
Homework - all current homework to turn in Monday, May 11

1. You do not need to read Section 8.4 but today’s material is based on it.

2. Turn in Monday, May 11: 8.1, 2, 3, 5