- In the following, I have plotted a set of data points:

$$(y_i, x_i) \quad i = 1, 2, \dots n$$

- The best least-squares linear fit for y has been found. This can be visualized as fitting the best line through a scatter plot

$$\mathrm{SE}(a, b) = \frac{1}{n-1} \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$$

- This is the "least-squares fit (see EE241)" and is referred to as a linear regression in statistics

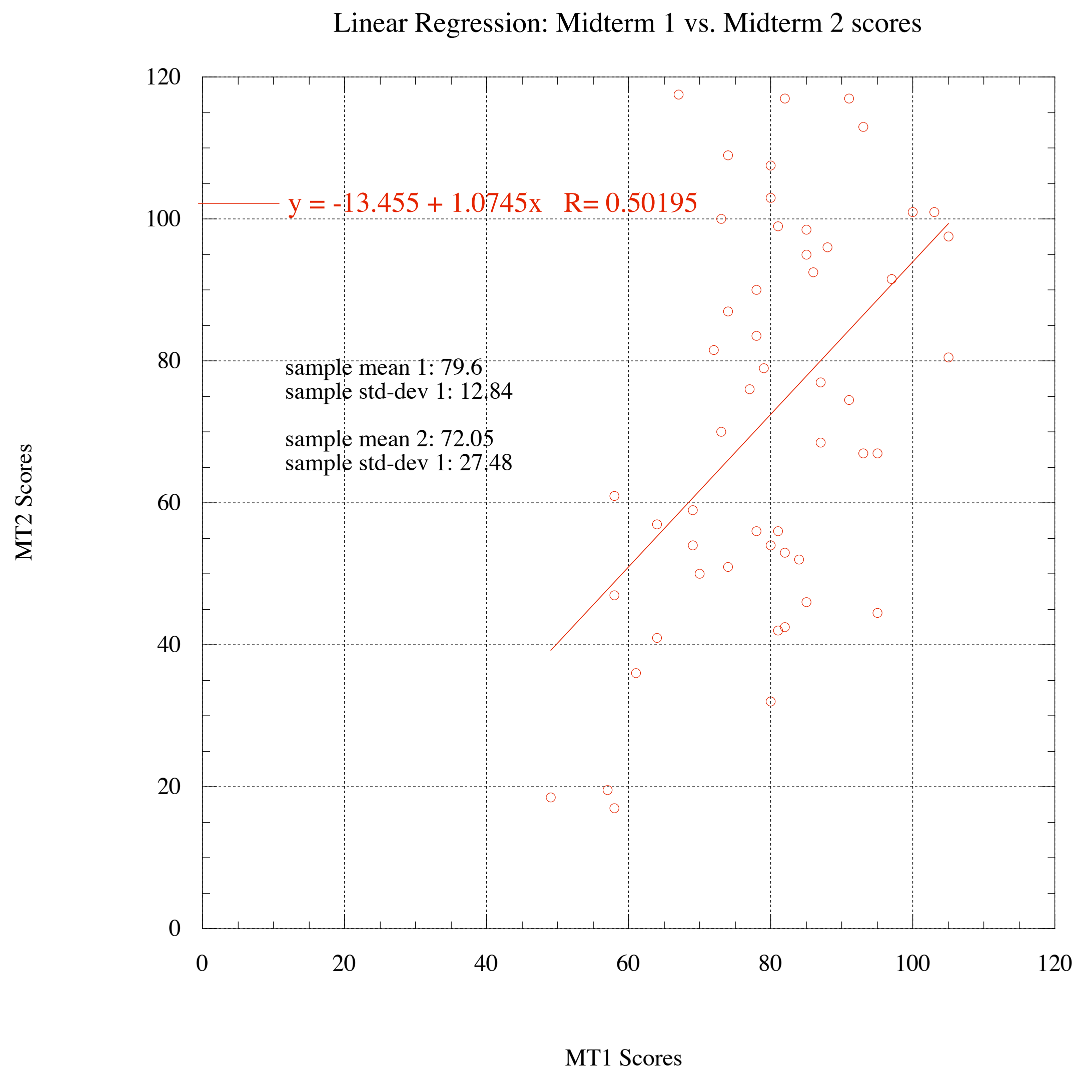
- The best choice for the coefficients can be shown to correspond to the LMMSE solution with the actual second moment quantities replaced by their corresponding sample quantities

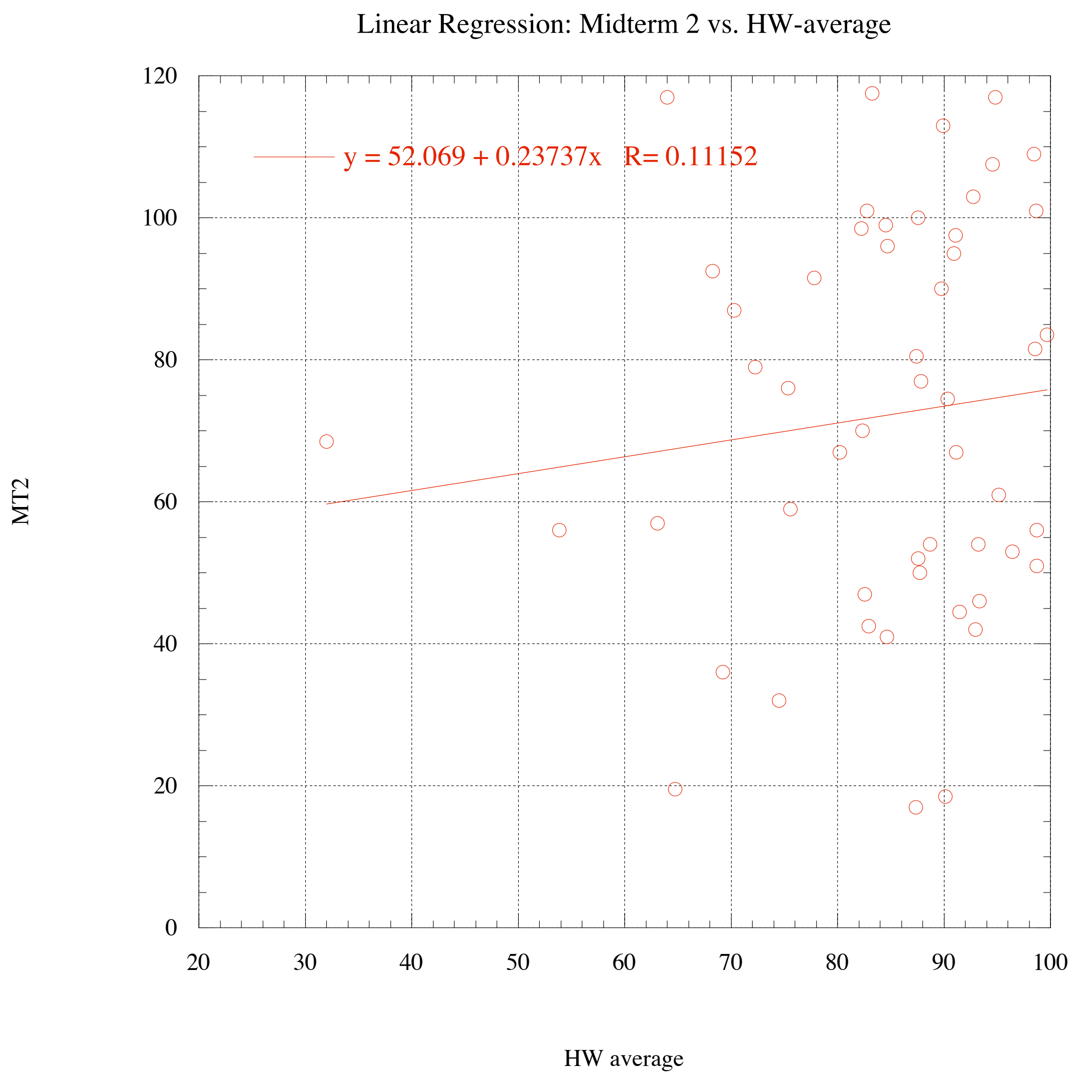
$$a_{\mathrm{opt}} = \frac{s_y}{s_x} r_{x,y} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$b_{\mathrm{opt}} = \bar{y} - a_{\mathrm{opt}} \bar{x}$$

$$r_{x,y} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{(n-1) s_x s_y}$$ (sample correlation coefficient)

$$(1 - r_{xy}^2) = \frac{\mathrm{SE}(a_{\mathrm{opt}}, b_{\mathrm{opt}})}{s_y^2}$$ is a measure of the "goodness of fit"


Linear Regression: Midterm 1 vs. Midterm 2 scores
y = -13.455 + 1.0745x R= 0.50195
sample mean 1: 79.6
sample std-dev 1: 12.84
sample mean 2: 72.05
sample std-dev 1: 27.48
MT2 Scores
MT1 Scores
0
20
40
60
80
100
120


Linear Regression: Midterm 2 vs. HW-average
y = 52.069 + 0.23737x R= 0.11152
120
100
80
60
40
20
0
20
30
40
50
60
70
80
90
100
MT2
HW average