Chapter 10, Review problem 5:

(a) *False:* Correlation quantifies the strength of the linear relationship between weight and lifting ability. It does not (by itself) provide a formula for predicting lifting ability from weight.

(b) *False:* This is just a reinterpretation of (a), and still false.

(c) *True:* This follows from the fact that the correlation is *positive* in this example.

(d) *True:* As above, this follows from the fact that $r > 0$ here.

(e) *False:* Not what the correlation tells us.[†]

Chapter 11, Review problem 5:

(a) We assume that the scores on the final exam (Avg= 55 and SD= 15) followed the normal curve (approximately) because the scatter plot is 'football shaped'. With this in mind, the percentage of students who scored 80 or more on the final is approximately equal to the area under the normal curve to the right of

$$z = \frac{80 - 55}{15} \approx 1.67$$

which is about 4.75% according to the normal table.

(b) The average score on the final exam of all students who scored 80 on the midterm is estimated by the regression line to be

$$\hat{y}(80) = 55 + 0.6 \times \frac{80 - 50}{25} \times 15 \approx 65.8,$$

and the standard deviation of all these scores is estimated by the R.M.S error of regression

$$\sqrt{1 - 0.6^2} \times 15 = 12.$$

Using the normal curve inside the vertical strip corresponding to a midterm score of 80, (Avg$\approx$ 65.8 and $SD \approx 12$), we find that the percentage of these students who scored over 80 on the final is approximately equal to the area under the normal curve to the right of

$$z = \frac{80 - 65.8}{12} \approx 1.18,$$

which is about 12.5% according to the table.

[†] Though there is a grain of truth to this one: in a certain sense, 36% ($= r^2$) of the variation in lifting ability can be explained by the subject's weight. This was not covered in class, and (d) is still false.

Chapter 12, Review problem 2:
The slope coefficient is

$$\beta_1 = r \cdot \frac{SD_{inc}}{SD_{ht}} = 0.2 \cdot \frac{20,000}{2.5} = 1600$$

and the intercept coefficient is

$$\beta_0 = \overline{inc} - \beta_1 \overline{ht} = 21000 - 1600 \cdot 64 = -81,400,$$

so the regression equation for predicting a woman's income from her height is

$$\widehat{inc}_j = -81400 + 1600ht_j.$$

Interpretation: The taller a woman is, the greater her income, on average. More precisely, we can say that for each additional inch in height, a woman is predicted to make \$1600 more per year on average. E.g., women who are 67 inches tall are predicted to make about \$3200 more per year than women who are 65 inches tall, on average.

Comment: Notice that the intercept coefficient in this example, which is negative and would correspond to the average income of women who are 0 inches tall, does *not* have a meaningful interpretation.