SUPPLEMENTARY EXERCISES FOR CHAPTER 8:
LEAST SQUARES EQUATION AND RELATED IDENTITIES

In these exercises you will derive information about the least squares equation and some identities among quantities involved in regression. Exercise 6 will be used in another handout to explain why r^2 is the fraction of the variation explained by regression.

Notation:

- We have values $x_1, x_2, \ldots, x_n$ for an explanatory variable x, and corresponding values $y_1, y_2, \ldots, y_n$ for the response variable y.
- The least squares line has equation $\hat{y} = b_0 + b_1x$.
- The predicted value for x_i is $\hat{y}_i = b_0 + b_1x_i$.
- The i^{th} residual is $e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1x_i)$.
- $\bar{x}$ = the mean of the x_i's, and $\bar{y}$ = the mean of the y_i's

Recall that the least squares line minimizes $\sum_{i=1}^{n} e_i^2$, the sum of the squares of the residuals. In other words, the coefficients b_0 and b_1 are chosen to minimize (as a function of b_0 and b_1) the function

$$f(b_0, b_1) = \sum_{i=1}^{n}[y_i - (b_0 + b_1x_i)]^2$$

From calculus, we know that for this to happen, (b_0, b_1) must be a critical point of f; that is, both partial derivatives of f (with respect to b_0 and with respect to b_1) must be zero.

Exercises

1) Find the partial derivatives of f with respect to b_0 and with respect to b_1 and set them equal to zero. The resulting equations are called the *normal equations*.

2) Use one of the normal equations to show that
   a) the residuals sum to zero (i.e., show that $\sum_{i=1}^{n} e_i = 0$), and hence that
   b) the mean of the residuals is zero. (This was stated without proof on p. 189 of the text.)

3) Use the other normal equation to establish the formula $\sum_{i=1}^{n} x_ie_i = 0$.

4) Show that the point $(\bar{x}, \bar{y})$ lies on the least squares line – i.e., that $\bar{y} = b_0 + b_i\bar{x}$. (This was state without proof on p. 182 of the text). [Hint: Note that $y_i = e_i + \hat{y}_i$ and use one of the above exercises.]

5) Show that the mean of the $\hat{y}_i$'s is $\bar{y}$.

6) Show that $\hat{y}_i - \bar{y} = b_1(x_i - \bar{x})$. [Hint: Exercise 4]

7) Show that $\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0$. [Hint: Start by using Exercise 6 and the definition of e_i, then use two of the other exercises.]