Statistics 252 (March 11, 2005)

Linear Regression and Least Squares

Consider the linear regression model $Y = \beta_0 + \beta_1 x + \varepsilon$ where ε is a mean zero random variable. Our goal is to predict the linear trend

$$\mathbb{E}(Y) = \beta_0 + \beta_1 x$$

by estimating the intercept and the slope of this line. That is, we seek estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ such that

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Our goal is to choose $\hat{\beta}_0$ and $\hat{\beta}_1$ in such a way that they minimize SSE, the sum of the squares of the errors. That is, we want to minimize

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where y_i is the ith observation of the random variable Y, and corresponds to the input x_i. Substituting the value of $\hat{y}$ into the equation for SSE we see that SSE can be viewed as a function of $\hat{\beta}_0$ and $\hat{\beta}_1$. We can now use the methods of elementary calculus to minimize this function. Namely, we find the first derivative, set it equal to 0, and solve for the critical points. We then use the second derivative test to check that the critical points are indeed minimizers. Thus,

$$\mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

Next, we find

$$\begin{aligned} \frac{\partial}{\partial \hat{\beta}_0} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= -2 \sum_{i=1}^{n} y_i + 2n\hat{\beta}_0 + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i \\ &= -2n\overline{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1 \overline{x} \end{aligned}$$

and

$$\frac{\partial}{\partial \hat{\beta}_1} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = -2 \sum_{i=1}^{n} x_i y_i + 2\hat{\beta}_0 \sum_{i=1}^{n} x_i + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i^2.$$

From the first equation,

$$\frac{\partial}{\partial \hat{\beta}_0} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

implies

$$-2n\overline{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1 \overline{x} = 0$$

so that

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}.$$

From the second equation,

$$\frac{\partial}{\partial \hat{\beta}_1} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

implies
$$-2\sum_{i=1}^{n} x_i y_i + 2\hat{\beta}_0 \sum_{i=1}^{n} x_i + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$
so that
$$-\sum_{i=1}^{n} x_i y_i + (\overline{y} - \hat{\beta}_1 \overline{x}) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0.$$
Distributing, and collecting $\hat{\beta}_1$ gives
$$-\sum_{i=1}^{n} x_i y_i + \overline{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \overline{x} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$
so that
$$\hat{\beta}_1 \left(\overline{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2 \right) = \overline{y} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i.$$
Thus, we find
$$\hat{\beta}_1 = \frac{\overline{y} \sum\limits_{i=1}^{n} x_i - \sum\limits_{i=1}^{n} x_i y_i}{\overline{x} \sum\limits_{i=1}^{n} x_i - \sum\limits_{i=1}^{n} x_i^2}.$$
Now, we can write this in a nicer way:
$$\hat{\beta}_1 = \frac{n\overline{y}\,\overline{x} - \sum\limits_{i=1}^{n} x_i y_i}{n\overline{x}^2 - \sum\limits_{i=1}^{n} x_i^2} = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{y}\,\overline{x}}{\sum\limits_{i=1}^{n} x_i^2 - n\overline{x}^2}.$$
With a bit of algebra, we can write the numerator as
$$\sum_{i=1}^{n} x_i y_i - n\overline{y}\,\overline{x} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) =: S_{XY}$$
and the denominator as
$$\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 =: S_{XX}.$$
Thus, we can write $\hat{\beta}_1$ as
$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}.$$
This version might be useful for remembering the formula, and it often appears when you are doing numerical calculations with computer software (for example, with SAS). The software tends to return the values of S_{XY} and S_{XX} because they are useful for *residual analysis* in general (this is not Stat 252 stuff). One final note: the formulæ for $\hat{\beta}_0$ and $\hat{\beta}_1$ agree with those in the Stat 151 textbook (pages 529–532 in the first edition and pages 491–494 in the second edition).

For Culture: The Multivariable Second Derivative Test

The second derivative test for functions of two variables is not required for Stat 252, although the ability to compute partial derivatives *IS* required for Stat 252.

Suppose that f is a (continuously differentiable) function of two variables, say $f(x, y)$. The *critical points* of the function f occur at those points (x_0, y_0) where (simultaneously)

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

The point (x_0, y_0) is a *local minimum* if

$$\frac{\partial^2 f}{\partial x^2}(x_0, y_0) > 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial x^2}(x_0, y_0) \cdot \frac{\partial^2 f}{\partial y^2}(x_0, y_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)\right)^2 > 0.$$

The point (x_0, y_0) is a *local maximum* if

$$\frac{\partial^2 f}{\partial x^2}(x_0, y_0) < 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial x^2}(x_0, y_0) \cdot \frac{\partial^2 f}{\partial y^2}(x_0, y_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)\right)^2 > 0.$$

[Yes, the inequalities are correct.]

Using this second derivative test, you can verify that $\hat{\beta}_0$ and $\hat{\beta}_1$ are, in fact, the minimizers of SSE. To quote our textbook, “We leave this for you to prove.” (page 538)

Exercise: If you are taking Math 212, Math 213, or Math 214 this semester, guess what?