From Wackerly DD, Mendenhall W and Scheaffer RL (2008) MATHEMATICAL STATISTICS WITH APPLICATIONS 7th Edn, Cengage, pp.569-571

11.3 The Method of Least Squares

A procedure for estimating the parameters of any linear model—the method of least squares—can be illustrated simply by fitting a straight line to a set of data points. Suppose that we wish to fit the model

$$E(Y) = \beta_0 + \beta_1 x$$

to the set of data points shown in Figure 11.5. [The independent variable x could be w^2 or $(w)^{1/2}$ or $\ln w$, and so on, for some other independent variable w.] That is, we postulate that $Y = \beta_0 + \beta_1 x + \varepsilon$, where ε possesses some probability distribution with $E(\varepsilon) = 0$. If $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators of the parameters β_0 and β_1, then $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$ is clearly an estimator of $E(Y)$.

The least-squares procedure for fitting a line through a set of n data points is similar to the method that we might use if we fit a line by eye; that is, we want the differences between the observed values and corresponding points on the fitted line to be "small" in some overall sense. A convenient way to accomplish this, and one that yields estimators with good properties, is to minimize the sum of squares of the vertical deviations from the fitted line (see the deviations indicated in Figure 11.5). Thus, if

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

is the predicted value of the ith y value (when $x = x_i$), then the deviation (sometimes called the *error*) of the observed value of y_i from $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is the difference $y_i - \hat{y}_i$ and the sum of squares of deviations to be minimized is

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2.$$

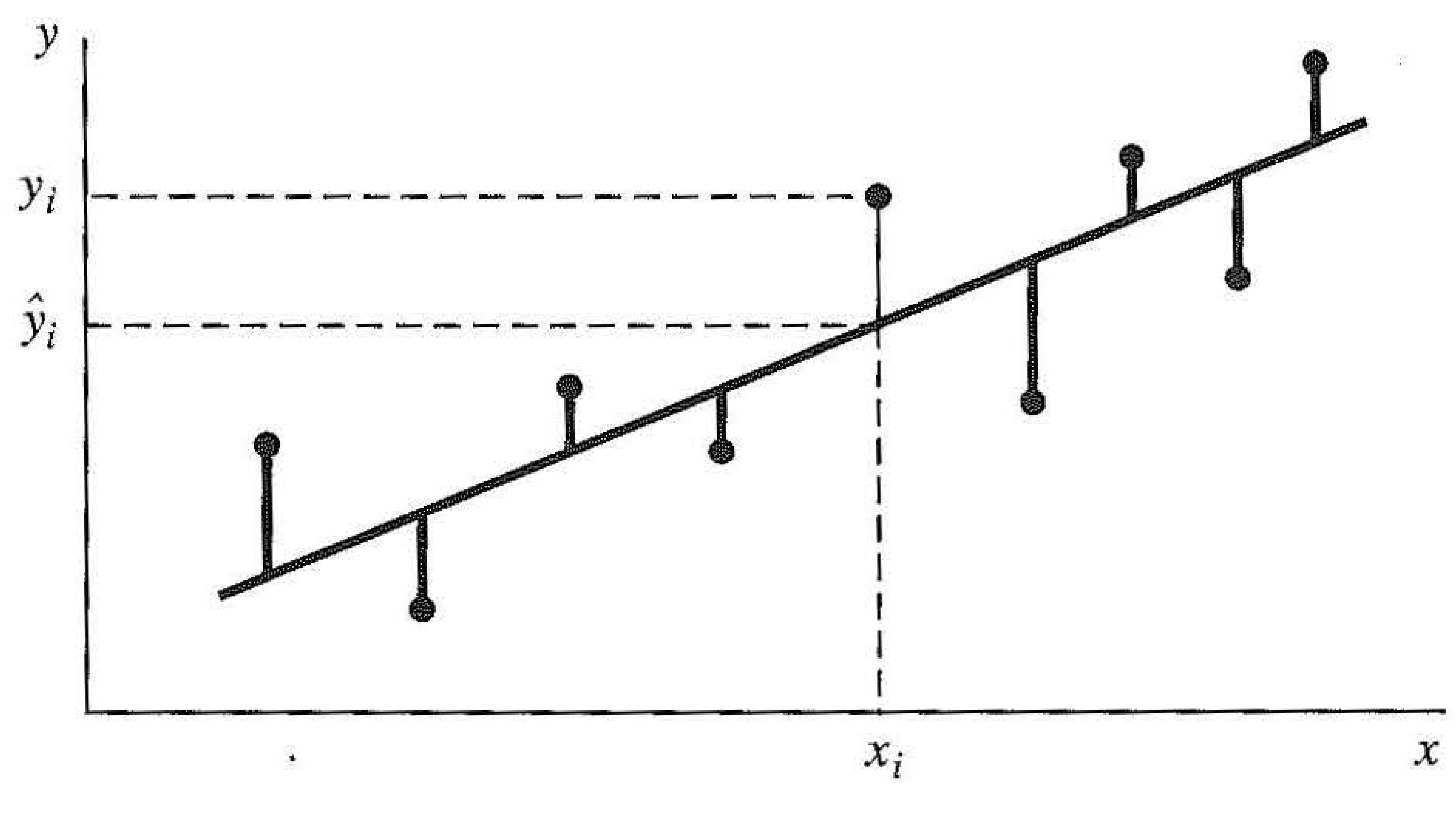

FIGURE 11.5 Fitting a straight line through a set of data points

The quantity SSE is also called the *sum of squares for error* for reasons that will subsequently become apparent.

If SSE possesses a minimum, it will occur for values of β_0 and β_1 that satisfy the equations, $\partial\text{SSE}/\partial\hat{\beta}_0 = 0$ and $\partial\text{SSE}/\partial\hat{\beta}_1 = 0$. Taking the partial derivatives of SSE with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and setting them equal to zero, we obtain

$$\frac{\partial\text{SSE}}{\partial\hat{\beta}_0} = \frac{\partial\left\{\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\right\}}{\partial\hat{\beta}_0} = -\sum_{i=1}^{n} 2[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]$$

$$= -2\left(\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i\right) = 0$$

and

$$\frac{\partial\text{SSE}}{\partial\hat{\beta}_1} = \frac{\partial\left\{\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2\right\}}{\partial\hat{\beta}_1} = -\sum_{i=1}^{n} 2[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]x_i$$

$$= -2\left(\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2\right) = 0.$$

The equations $\partial\text{SSE}/\hat{\beta}_0 = 0$ and $\partial\text{SSE}/\hat{\beta}_1 = 0$ are called the *least-squares equations* for estimating the parameters of a line.

The least-squares equations are linear in $\hat{\beta}_0$ and $\hat{\beta}_1$ and hence can be solved simultaneously. You can verify that the solutions are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2},$$

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

Further, it can be shown that the simultaneous solution for the two least-squares equations yields values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize SSE. We leave this for you to prove.

The expressions

$$\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \quad \text{and} \quad \sum_{i=1}^{n}(x_i - \overline{x})^2$$

that are used to calculate $\hat{\beta}_1$ are often encountered in the development of simple linear regression models. The first of these is calculated by summing products of x-values minus their mean and y-values minus their mean. In all subsequent discussions, we will denote this quantity by S_{xy}. Similarly, we will denote the second quantity by S_{xx} because it is calculated by summing products that involve only the x-values.

Least-Squares Estimators for the Simple Linear Regression Model

1. $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$, where $S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$ and $S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2$.
2. $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$.