## 8.5.3 The Method of Least Squares

Here, we use a different method to estimate $\beta_0$ and $\beta_1$. This method will result in the same estimates as before; however, it is based on a different idea. Suppose that we have data points $(x_1, y_1)$, $(x_2, y_2)$, $\cdots$, $(x_n, y_n)$. Consider the model

$$\hat{y} = \beta_0 + \beta_1 x.$$

The errors (residuals) are given by

$$e_i = y_i - \hat{y}_i = y_i - \beta_0 - \beta_1 x.$$

The *sum of the squared errors* is given by

$$g(\beta_0, \beta_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \qquad (8.7)$$

To find the best fit for the data, we find the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ such that $g(\beta_0, \beta_1)$ is minimized. This can be done by taking partial derivatives with respect to $\beta_0$ and $\beta_1$, and setting them to zero. We obtain

$$\frac{\partial g}{\partial \beta_0} = \sum_{i=1}^{n} 2(-1)(y_i - \beta_0 - \beta_1 x_i) = 0, \qquad (8.8)$$
$$\frac{\partial g}{\partial \beta_1} = \sum_{i=1}^{n} 2(-x_i)(y_i - \beta_0 - \beta_1 x_i) = 0. \qquad (8.9)$$

By solving the above equations, we obtain the same values of $\hat{\beta}_0$ and $\hat{\beta}_1$ as before

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}},$$
$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x},$$

where

$$s_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2,$$
$$s_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}).$$

This method is called the method of **least squares**, and for this reason, we call the above values of $\hat{\beta}_0$ and $\hat{\beta}_1$ the **least squares estimates** of $\beta_0$ and $\beta_1$.