2.1.2 Least Squares Fitting

We select the best values of β_0 and β_1 by minimizing the *error in fit*. For two data points (x_1, y_1) and (x_2, y_2), the errors in fit are

$$
\begin{aligned}
e_1 &= y_1 - (\beta_0 + \beta_1 x_1) \\
e_2 &= y_2 - (\beta_0 + \beta_1 x_2)
\end{aligned}
$$

respectively. But note that, potentially, $e_1 > 0$ and $e_2 < 0$ so there is a possibility that these fitting errors cancel each other out. Therefore we look at **squared** errors (as a large negative error is as bad as a large positive error)

$$
\begin{aligned}
e_1^2 &= (y_1 - (\beta_0 + \beta_1 x_1))^2 \\
e_2^2 &= (y_2 - (\beta_0 + \beta_1 x_2))^2
\end{aligned}
$$

For n data, we obtain n misfit squared errors

$$e_1^2, \ldots, e_n^2$$

We select β_0 and β_1 as the values of the parameters that minimize SSE, where

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

We wish to make the total misfit squared error as small as possible.

SSE - sum of squared errors - is similar to the SSE for ANOVA. We could write

$$SSE = SSE(\beta_0, \beta_1)$$

to show the dependence of SSE on the parameters.

Minimization of $SSE(\beta_0, \beta_1)$ is achieved **analytically**.

Two routes: (i) calculus and (ii) geometric methods. It follows that the best parameters $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are given by

$$\widehat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} \qquad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

where

- Sum of Squares SS_{xx}:

$$SS_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

- Sum of Squares SS_{xy}:

$$SS_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

$\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the **least-squares estimates**

$$y = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

is the **least-squares line of best fit**. The **fitted-values** are

$$\hat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \qquad i = 1, \ldots, n$$

and the **residuals** or **residual errors** are

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \qquad i = 1, \ldots, n$$

2.1.3 Model Assumptions for Least-Squares

To utilize least-squares for the probabilistic model

$$Y = \beta_0 + \beta_1 x + \epsilon$$

we make the following assumptions

1. The expected error $E[\epsilon]$ is zero so that

$$E[Y] = \beta_0 + \beta_1 x$$

2. The variance of the error, $Var[\epsilon]$, is constant and does not depend on x.
3. The probability distribution of ϵ is a symmetric distribution about zero (a stronger assumption is that ϵ is Normally distributed).
4. The errors for two different measured responses are independent, i.e. the error ϵ_1 in measuring y_1 at x_1 is independent of the error ϵ_2 in measuring y_2 at x_2.

2.1.4 Parameter Estimation: Estimating σ^2

Using the LS procedure, we can construct an estimate of the *error* or *residual error* variance

Recall that

$$Var[\epsilon] = \sigma^2$$

An estimate of σ^2 is

$$\widehat{\sigma}^2 = \frac{SSE(\widehat{\beta}_0, \widehat{\beta}_1)}{n-2} = s^2$$

say.

Note that the denominator $n-2$ is again a *degrees of freedom* parameter of the form

$$\text{TOTAL NUMBER OF DATA} \; - \; \text{NUMBER OF PARAMETERS ESTIMATED}$$

or $n-p$, where in the simple linear regression, $p=2$ ($\widehat{\beta}_0$ and $\widehat{\beta}_1$). Note also that

$$SSE(\widehat{\beta}_0, \widehat{\beta}_1) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = SS_{yy} - \widehat{\beta}_1 SS_{xy}$$

where

$$SS_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2$$