Proof that $\mathsf{E}(RSS) = (n-2)\sigma^2$

Recall from the notes that the residual sum of squares, RSS, can be written as:

$$S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

and that

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

Now note that these can be put together to give

$$RSS = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = S_{yy} - \left(\frac{S_{xy}^2}{S_{xx}^2}\right) S_{xx} = S_{yy} - \hat{\beta}^2 S_{xx}$$

The expectation of this can therefore be found by finding in turn the expectations of S_{yy} and $\hat{\beta}^2 S_{xx}$. The second of these is the simpler. We know from the notes that:

$$\begin{aligned} \mathsf{E}(\hat{\beta}) &= \beta \\ \mathsf{var}(\hat{\beta}) &= \frac{\sigma^2}{S_{xx}} \end{aligned}$$

We also know from the definition of the variance of a random variable, X, that:

$$\mathsf{E}(X^2) = \mathsf{var}(X) + \mathsf{E}(X)^2$$

Putting these together shows that:

$$\begin{aligned} \mathsf{E}(\hat{\beta}^2 S_{xx}) &= \left(\mathsf{var}(\hat{\beta}) + \mathsf{E}(\hat{\beta})^2\right) S_{xx} \\ &= \left(\frac{\sigma^2}{S_{xx}} + \beta^2\right) S_{xx} \\ &= \sigma^2 + \beta^2 S_{xx} \qquad (1) \end{aligned}$$

The expectation of $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$ might be thought to be $(n-1)\sigma^2$, as the expression is formally the same as that for a sample variance. However, this result arises only for the case when the y_i have the same distribution and in the case of the regression model the expectations of the y_i are different for observations with different x_is. How do we proceed? The ϵ_i do have a common distribution and so we know from the case of the variance of a simple random sample that the expectation of $\sum_{i=1}^n (\epsilon_i - \bar{\epsilon})^2$ is $(n-1)\sigma^2$. Can we make use of this observation when working out $\mathsf{E}(S_{yy})$?

As $y_i = \alpha + \beta x_i + \epsilon_i$, then summing this from $i = 1, \ldots, n$ and dividing by n shows that $\bar{y} = \alpha + \beta\bar{x} + \bar{\epsilon}$. Therefore

$$\begin{aligned} y_i - \bar{y} &= (\alpha + \beta x_i + \epsilon_i) - (\alpha + \beta\bar{x} + \bar{\epsilon}) \\ &= \beta(x_i - \bar{x}) + (\epsilon_i - \bar{\epsilon}) \end{aligned}$$

Squaring this gives:

$$(y_i - \bar{y})^2 = \beta^2(x_i - \bar{x})^2 + 2\beta(x_i - \bar{x})(\epsilon_i - \bar{\epsilon}) + (\epsilon_i - \bar{\epsilon})^2$$

and summing from $i = 1, \ldots, n$ gives:

$$S_{yy} = \beta^2 S_{xx} + 2\beta\sum_{i=1}^{n}(x_i - \bar{x})(\epsilon_i - \bar{\epsilon}) + \sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2$$

Since $\mathsf{E}(\epsilon_i) = 0$ it follows that $\mathsf{E}(\bar{\epsilon}) = 0$ and hence the expectation of the second term on the RHS is 0. Therefore

$$\begin{aligned} \mathsf{E}(S_{yy}) &= \beta^2 S_{xx} + \mathsf{E}\left(\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2\right) \\ &= \beta^2 S_{xx} + (n-1)\sigma^2 \end{aligned} \tag{2}$$

Now subtract (1) from (2) to get

$$\mathsf{E}(RSS) = (n-2)\sigma^2$$

as required.