Section 1.7. Estimation of Error Terms Variance σ^2

Note. In this section we estimate the variance σ^2 of the error terms ε_i of the regression model. In Chapter 2 we will use these estimates in inferences concerning the regression function; in particular, see Section 2.6. Confidence Band for Regression Line.

Note. If Y_i are observations from a single population (or are all based on a single value of the predictor variable of X), then the estimated mean $\overline{Y}$ determines the *sample variance* s^2 as $s^2 = \dfrac{\sum_{i=1}^n (Y_i - \overline{Y})^2}{n-1}$. That is, it is the square of the "deviations" of the Y_i about their estimated mean $\overline{Y}$ divided by the degrees of freedom. Kutner et al. state (see page 25) that the degrees of freedom are $n-1$ "because one degree of freedom is lost by using $\overline{Y}$ as an estimate of the unknown population mean μ." Informally, when we have n measurements and compute a population statistic, the degrees of freedom are n minus the number of parameters that are estimated in the computation of the statistic. So the deviation of observation Y_i is based on $\hat{Y}_i$: $Y_i - \hat{Y}_i = e_i$.

Definition. For data points (X_i, Y_i) where $i = 1, 2, \ldots, n$, the *error sum of squares* or *residual sum of squares*, denoted SSE, is

$$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2.$$

The *error mean square* or *residual mean square*, denoted MSE, is

$$s^2 = MSE = \frac{SSE}{n-2} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum_{i=1}^n e_i^2}{n-2}.$$

Note 1.7.A. There are $n-2$ degrees of freedom since the regression model implies $\hat{Y}_i = b_0 + b_1 X_i$ where b_0 and b_1 are estimates of β_0 and β_1, respectively; that is, two parameters are estimated in the computation of MSE. An estimator of the standard deviation σ is simply $s = \sqrt{MSE}$.

Example 1.7.A. In Example 1.6.C, based on the Depth and Age data of Example 1.6.A, we see for the given data that the sum of the squares of the residuals squared is $SSE = \sum_{i=1}^{n} 242754.49$ years^2. Since $n = 7$ then there are $n - 2 = 5$ degrees of freedom and the error mean square is

$$s^2 = MSE = \frac{MSE}{n-2} = \frac{242754.49}{5} = 48550.90 \text{ years}^2.$$

So the estimator of the standard deviation σ is

$$s = \sqrt{MSE} = \sqrt{48550.90} = 220.34 \text{ years}.$$

In Example 1.6.B, we saw that when depth $X = 600$ cm, $\hat{Y} = 7704.74$ years. Since $s = 220.34$ years, we can use this information to find confidence intervals for the value of Y when $X = 600$ cm, as we'll explore later (see Section 2.4. Interval Estimation of $E\{Y_h\}$).

Revised: 9/22/2022