UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS

Generalised Regression Models

GRM: Example — Quadratic regression **Semester 1, 2022–2023**

The file Adhesive.txt contains the data presented in Numerical Example 1.7 on the effect of setting time on the setting strength of a woodwork adhesive. The two columns contain the forces, y, (in kg) required to separate two strips of wood and the setting times, x, (in hours) for which the strips were clamped together. The separating force appears to increase with setting time and then decrease, and it looks as if a quadratic regression of force on setting time would fit the data well. If we define a new explanatory variable to be the square of setting time, then the vector of responses and the matrix of the explanatory variables are given by

$$\mathbf{y}^T = \begin{pmatrix} 35.1 & 39.7 & 40.3 & 42.0 & 44.1 & 47.0 & 46.0 & 48.1 & 49.9 & 44.3 & 45.1 & 46.2 \end{pmatrix}$$

$$\mathbf{X}^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 12 & 12 & 12 & 24 & 24 & 24 & 36 & 36 & 36 & 48 & 48 & 48 \\ 144 & 144 & 144 & 576 & 576 & 576 & 1296 & 1296 & 1296 & 2304 & 2304 & 2304 \end{pmatrix}$$

The matrix of (uncorrected) sums of squares and products of the explanatory variables and the vector of sums of products with the response are

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 12 & 360 & 12960 \\ 360 & 12960 & 518400 \\ 12960 & 518400 & 22021632 \end{pmatrix}, \quad \mathbf{X}^T\mathbf{y} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{pmatrix} = \begin{pmatrix} 527.8 \\ 16268.4 \\ 592286.4 \end{pmatrix}.$$

The inverse of $\mathbf{X}^T\mathbf{X}$ and the vector of least squares estimates are as follows

$$\left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 2.583 & -0.1875 & 0.00289 \\ -0.1875 & 0.01493 & -0.000241 \\ 0.00289 & -0.000241 & 0.00000402 \end{pmatrix}, \quad \widehat{\beta} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} = \begin{pmatrix} 26.950000 \\ 1.117778 \\ -0.015278 \end{pmatrix}.$$

The vectors of fitted values and residuals are given by

$$\left(\mathbf{X}\widehat{\beta}\right)^T = \begin{pmatrix} 38.16 & 38.16 & 38.16 & 44.98 & 44.98 & 44.98 & 47.39 & 47.39 & 47.39 & 45.40 & 45.40 & 45.40 \end{pmatrix}$$

$$\mathbf{e}^T = \begin{pmatrix} -3.06 & 1.54 & 2.14 & -2.98 & -0.88 & 2.02 & -1.39 & 0.71 & 2.51 & -1.10 & -0.30 & 0.80 \end{pmatrix}$$

The (uncorrected) sum of squares of the responses, the sum of squares for fitting the quadratic model and the residual sum of squares are respectively

$$\begin{aligned} \mathbf{y}^T\mathbf{y} &= \sum_i y_i^2 \\ &= 23400.56\,, \\ \widehat{\beta}^T\mathbf{X}^T\mathbf{y} &= 26.950 \times 527.8 + 1.117778 \times 16268.4 - 0.015278 \times 592286.4 \\ &= 23359.7\,, \\ \mathbf{y}^T\mathbf{y} - \widehat{\beta}^T\mathbf{X}^T\mathbf{y} &= 23400.56 - 23359.7 \\ &= 40.9\,. \end{aligned}$$

Hence the residual mean square is $\widehat{\sigma}^2 = \frac{40.9}{9} = 4.54$, and the variances and covariances of the estimators of the regression coefficients are estimated by

$$\widehat{\sigma}^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \begin{pmatrix} 11.74 & -0.852 & 0.0131 \\ -0.852 & 0.0678 & -0.00110 \\ 0.0131 & -0.00110 & 0.0000183 \end{pmatrix}.$$

Analysis using the alternative model

The calculations are easier (and more accurate) if *corrected* sums of squares and products are used, i.e. if x_i and x_i^2 are expressed as deviations from their means (which are 30 and 1080). The matrix of corrected sums of squares and products of the two explanatory variables and the vector of corrected products with the response are

$$\dot{\mathbf{X}}^T\dot{\mathbf{X}} = \begin{pmatrix} 12960 - \frac{360^2}{12} & 518400 - \frac{360\times 12960}{12} \\ 518400 - \frac{360\times 12960}{12} & 22021632 - \frac{12960^2}{12} \end{pmatrix} = \begin{pmatrix} 2160 & 129600 \\ 129600 & 8024832 \end{pmatrix},$$

$$\dot{\mathbf{X}}^T\mathbf{y} = \begin{pmatrix} 16268.4 - \frac{360\times 527.8}{12} \\ 592286.4 - \frac{12960\times 527.8}{12} \end{pmatrix} = \begin{pmatrix} 434.4 \\ 22262.4 \end{pmatrix}.$$

The inverse of $\dot{\mathbf{X}}^T\dot{\mathbf{X}}$ and the least squares estimates are then given by

$$\left(\dot{\mathbf{X}}^T\dot{\mathbf{X}}\right)^{-1} = \begin{pmatrix} 0.01493 & -0.000241 \\ -0.000241 & 0.00000402 \end{pmatrix},$$

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} = \left(\dot{\mathbf{X}}^T\dot{\mathbf{X}}\right)^{-1}\dot{\mathbf{X}}^T\mathbf{y} = \begin{pmatrix} 1.117778 \\ -0.015278 \end{pmatrix},$$

$$\widehat{\beta}_0 = \frac{527.8}{12} - 30\times 1.117778 - 1080\times(-0.015278) = 26.950.$$

The residual sum of squares is calculated as the total sum of squares *about the mean* minus the regression sum of squares:

$$\begin{aligned} S_{yy} - \widehat{\beta}_1 S_{1y} - \widehat{\beta}_2 S_{2y} &= \left(23400.56 - \frac{527.8^2}{12}\right) - 1.117778\times 434.4 - (-0.015278)\times 22262.4 \\ &= 186.157 - 145.443 \quad \text{(using more d.ps. than shown)} \\ &= 40.71. \end{aligned}$$

The residual sum of squares differs slightly from the value calculated previously but this is simply due to rounding errors in the calculation in the previous analysis.
Note that $\left(\dot{\mathbf{X}}^T\dot{\mathbf{X}}\right)^{-1}$ is equal to the lower right 2×2 sub-matrix of $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, so that the variances and covariance of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are identical to those of the previous analysis.