Level M Regression Models Examples

Writing models in vector-matrix notation

Consider the following model:

Data: $(y_i, x_i) \quad i = 1, \ldots, n$
Model: $E(Y_i) = \alpha + \beta x_i + \gamma x_i^2, Var(Y_i) = \sigma^2$
This model has been proposed for the potato example used in the lectures.

a) Write this model in vector-matrix form, $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, clearly identifying the elements of $\mathbf{Y}$, $\mathbf{X}$ and $\boldsymbol{\beta}$.

b) Identify formulae (in terms of the y_i's and x_i's) for the elements of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}^T\mathbf{Y}$.

Solution

a) $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ . \\ . \\ . \\ . \\ . \\ . \\ y_n \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & x_n & x_n^2 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$$

b)

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} n & \sum_i x_i & \sum_i x_i^2 \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i^3 \\ \sum_i x_i^2 & \sum_i x_i^3 & \sum_i x_i^4 \end{pmatrix} \quad \mathbf{X}^T\mathbf{Y} = \begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \\ \sum_i x_i^2 y_i \end{pmatrix}$$

Fitting the X-ray data

Let us consider the X-ray data again

Data: $(y_i, x_i), \quad i = 1, \ldots, n; \quad n = 15.$
$y_i = \log(N(t)/N(0)),$
x_i = time of irradiation.
Possible model: $y_i = \beta x_i + \epsilon_i$ for some β.

a) Estimate $\hat{\beta}$ using the summary statistics

$$\sum_{i=1}^{15} x_i^2 = 44640, \sum_{i=1}^{15} x_i y_i = -1638.263.$$

Solution

$$\begin{aligned}\hat{\beta} &= (X^TX)^{-1}X^TY \\ &= \left(\sum_{i=1}^{15} x_i^2\right)^{-1} \sum_{i=1}^{15} x_i y_i \\ &= \frac{-1638.263}{44640} \\ &= -0.0367\end{aligned}$$

Writing linear model for Drug Carbinazole data

In a study of the drug Carbinazole in the treatment of thyrotoxicosis, there is interest in comparing patients who are responders (R) to the drug with those who are dormant (D), i.e. non-responders. The patients are compared using their Protein Bound Iodine (PBI) levels.

Data from 20 responders and 20 dormants are available. The model for these data can be written as:

Data: $y_{ij}; i = 1, 2; j = 1, ..., 20$

Model: $E(Y_{1j}) = \theta_R$, $E(Y_{2j}) = \theta_D$, $\quad Var(Y_{ij}) = \sigma^2$

i.e. $E(Y_{ij}) = x_{1j}\theta_R + x_{2j}\theta_D$ where $(x_{1j}, x_{2j}) \begin{array}{ll} = (1,0), & \text{for } i = 1 \\ = (0,1), & \text{for } i = 2 \end{array}$

a) Write this model in vector matrix notation

b) Estimate $\hat{\theta_R}$ and $\hat{\theta_D}$ using these summary statistics

$$\sum_{j=1}^{20} y_{1j} = 196.7, \sum_{j=1}^{20} y_{2j} = 271.9, \sum_{j=1}^{20} y_{ij}^2 = 5816.20$$

c) Estimate RSS.

Solution

In vector-matrix form, this model is represented as, $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$:

$$E\begin{pmatrix} y_{1,1} \\ . \\ y_{1,20} \\ y_{2,1} \\ . \\ y_{2,20} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ . & . \\ 1 & 0 \\ 0 & 1 \\ . & . \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_R \\ \theta_D \end{pmatrix}$$

$$\sum_{j=1}^{20} y_{1j} = 196.7, \sum_{j=1}^{20} y_{2j} = 271.9, \sum_{j=1}^{20} y_{ij}^2 = 5816.20$$

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 20 & 0 \\ 0 & 20 \end{pmatrix} \quad (\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{20}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \mathbf{X}^T\mathbf{Y} = \begin{pmatrix} \sum_j y_{1j} \\ \sum_j y_{2j} \end{pmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \begin{pmatrix} \sum_j y_{1j}/20 \\ \sum_j y_{2j}/20 \end{pmatrix} = \begin{pmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \end{pmatrix} = \begin{pmatrix} 9.835 \\ 13.595 \end{pmatrix}$$

$$\begin{aligned} RSS = \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X}\hat{\boldsymbol{\beta}} = & \sum_i \sum_j y_{ij}^2 - \begin{pmatrix} \sum_j y_{1j} & \sum_j y_{2j} \end{pmatrix} \begin{pmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \end{pmatrix} \\ = & \ 5816.20 - (196.7 \ \ 271.9) \begin{pmatrix} 9.835 \\ 13.595 \end{pmatrix} = 185.175 \end{aligned}$$