

Multiple Linear Regression

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

January 31, 2008

Multiple Linear Regression

- Most interesting questions in biology involve relationships between multiple variables.
- There are typically multiple explanatory variables.
- Interactions between variables can be important in understanding a process.
- We will now study statistical models for when there is a single continuous quantitative response variable and multiple explanatory variables.
- Explanatory variables may be quantitative or factors (categorical variables).

An Example

- We will consider a small subset of the FEV data set.
- There are $n = 6$ children for whom we will develop a model to predict *forced expiratory volume* on the basis of *age*, *height*, and *sex* with a linear model.
- Here is the data for the example.

```

fev  age  ht   sex
1.72  7  54.5 female
1.74  8  54.0 male
2.09  9  59.5 male
3.13 10  62.0 male
2.87 11  60.5 female
2.57 12  63.0 female

```

Model Matrix

- The inputs are represented by a matrix of *predictors*.
- The intercept corresponds to a vector of ones.
- Each quantitative variable is a single column.
- Each categorical variable with m levels is represented by $m - 1$ columns.
- The i th row has all information about the i th individual in the sample.

$$y = \begin{pmatrix} 1.72 \\ 1.74 \\ 2.09 \\ 3.13 \\ 2.87 \\ 2.57 \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 7 & 54.5 & 0 \\ 1 & 8 & 54.0 & 1 \\ 1 & 9 & 59.5 & 1 \\ 1 & 10 & 62.0 & 1 \\ 1 & 11 & 60.5 & 0 \\ 1 & 12 & 63.0 & 0 \end{pmatrix}$$

Model Coefficients

- The parameters of a model are written as a vector β of size k .
- The i th row of X is denoted X_i .
- The model

$$y_i = \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + e_i$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$$

is written in matrix form as

$$y_i = X_i \beta + e_i.$$

where the $e_i \sim N(0, \sigma^2)$.

Model Coefficients

- In matrix multiplication, the dot product of the i th row of the matrix on the left times the j th column of the matrix on the right is the ij element of the product.
- The number of columns of the left matrix must match the number of rows of the right matrix.
- For example,

$$X_2 \beta = \begin{pmatrix} 1 & 8 & 54.0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$$

$$= 1 \cdot \beta_1 + 8 \cdot \beta_2 + 54.0 \cdot \beta_3 + 1 \cdot \beta_4$$

Least Squares

- The difference $y_i - X_i\beta$ measures the distance between the i th outcome and its prediction.
- In matrix notation, the sum of squared differences

$$\sum_{i=1}^n (y_i - X_i\beta)^2$$

is written

$$(y - X\beta)^t(y - X\beta).$$

- The t stands for *transpose* where the rows and columns of a matrix are swapped.
- X is an $n \times k$ matrix and β is a $k \times 1$ matrix (or vector) so the product is an $n \times 1$ matrix.
- The transpose turns an $n \times 1$ matrix into a $1 \times n$ matrix.
- The product of a $1 \times n$ matrix and a $n \times 1$ matrix is a 1×1 matrix, or a single number.

Example

$$y - X\beta = \begin{pmatrix} 1.72 - (1\beta_1 + 7\beta_2 + 54.5\beta_3 + 0\beta_4) \\ 1.74 - (1\beta_1 + 8\beta_2 + 54.0\beta_3 + 1\beta_4) \\ 2.09 - (1\beta_1 + 9\beta_2 + 59.5\beta_3 + 1\beta_4) \\ 3.13 - (1\beta_1 + 10\beta_2 + 62.0\beta_3 + 1\beta_4) \\ 2.87 - (1\beta_1 + 11\beta_2 + 60.5\beta_3 + 0\beta_4) \\ 2.57 - (1\beta_1 + 12\beta_2 + 63.0\beta_3 + 0\beta_4) \end{pmatrix}$$

$$\begin{aligned} (y - X\beta)^t(y - X\beta) &= (1.72 - (1\beta_1 + 7\beta_2 + 54.5\beta_3 + 0\beta_4))^2 \\ &+ \dots \\ &+ (2.57 - (1\beta_1 + 12\beta_2 + 63.0\beta_3 + 0\beta_4))^2 \end{aligned}$$

Least Squares

- The least squares criterion says that the best choice for β is the one where the sum of squared residuals, $(y - X\beta)^t(y - X\beta)$, is minimized.
- In theory, to find this we could take derivatives of $(y - X\beta)^t(y - X\beta)$ with respect to β_j , for $j = 1, \dots, k$, set each of these k equations to 0 and solve.
- In matrix notation, taking derivatives and doing some matrix algebra leads to the expression

$$X^t y = X^t X \beta$$

- Notice that each side of the equation is a $k \times 1$ vector.
- On the left, $(k \times n) \cdot (n \times 1) \rightarrow (k \times 1)$ and on the right, $(k \times n) \cdot (n \times k) \cdot (k \times 1) \rightarrow (k \times 1)$.

Matrix Inverses

- Notice that $X^t X$ is a $k \times k$ matrix.
- Some square matrices have *inverses*, square matrices of the same size where the product is the *identity matrix* I , a matrix with all zeros except for ones along the main diagonal.
- So, $AA^{-1} = A^{-1}A = I$ if A is a $k \times k$ matrix with an inverse.
- The identity matrix is special and acts like the number 1 — for any matrices A and B of the right dimension, $AI = A$ and $IB = B$.

Example

- In our example,

$$X^t X = \begin{pmatrix} 6 & 57 & 354 & 3 \\ 57 & 559 & 3390 & 27 \\ 354 & 3390 & 20900 & 176 \\ 3 & 27 & 176 & 3 \end{pmatrix}$$

$$(X^t X)^{-1} = \begin{pmatrix} 144.178 & 6.047 & -3.443 & 2.844 \\ 6.047 & 0.386 & -0.167 & 0.247 \\ -3.443 & -0.167 & 0.086 & -0.095 \\ 2.844 & 0.247 & -0.095 & 0.834 \end{pmatrix}$$

Least Squares Solution

- The equation

$$X^t y = X^t X \beta$$

is solved for β by

$$\hat{\beta} = (X^t X)^{-1} X^t y$$

- With our example,

$$\hat{\beta} = \begin{pmatrix} -5.285 \\ 0.022 \\ 0.126 \\ 0.060 \end{pmatrix}, \quad y = \begin{pmatrix} 1.72 \\ 1.74 \\ 2.09 \\ 3.13 \\ 2.87 \\ 2.57 \end{pmatrix}, \quad \text{and} \quad \hat{y} = X \hat{\beta} = \begin{pmatrix} 1.72 \\ 1.73 \\ 2.45 \\ 2.78 \\ 2.56 \\ 2.89 \end{pmatrix}$$

Geometry

- There is also a geometric interpretation of least squares regression.
- Each predictor, or column of X , is a vector in an n -dimensional space.
- These k vectors form a k -dimensional *hyper-plane* in this n -dimensional space.
- This hyper-plane represents all possible fitted values for a given set of predictors.
- The fitted values \hat{y} are as close as possible to the outcome vector y .
- \hat{y} is the projection of y into the hyper-plane.
- The residual vector $y - \hat{y}$ is orthogonal to every predictor.
- See the chalkboard picture!