3 Multiple Regression I

<u>Motivation</u>

Consider a study assessing the affect a drug has on someone's resting heart rate. The researcher randomly places the respondents into two groups; control group and drug group. She then measures each respondent's resting heart rate 1 hour after the drug was administered. The simple linear regression model for this experiment is

$$Y = \beta_0 + \beta_1 x + \epsilon, \quad x = \begin{cases} 1 & \text{control group} \\ 0 & \text{drug group} \end{cases}, \quad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

To investigate the resting heart rate between the two groups, she will run a t-test on the slope parameter β_1.

- What other variables should we include in the model?

Extended models:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \quad x_1 = \begin{cases} 1 & \text{control group} \\ 0 & \text{drug group} \end{cases}, \quad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon, \quad x_1 = \begin{cases} 1 & \text{control group} \\ 0 & \text{drug group} \end{cases}, \quad \epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

The **multiple linear regression** model:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i, \quad n = 1, \ldots, n, \tag{3.1}$$

with

$$\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

Note: There are p “beta" parameters ($\beta_0, \beta_1, \ldots, \beta_{p-1}$) and one dispersion parameter σ^2.

The **estimated** multiple linear regression model:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_{p-1} x_{p-1}. \tag{3.2}$$

Finding the least squares estimators of the above model requires some machinery.

3.1 Primer on Matrix Algebra

DEFINITION 3.1 A **matrix** is a rectangular array of numbers, symbols, or expressions, arranged in rows and columns. Typically matrices are denoted with capital bold letters.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{pmatrix}$$

a_{ij} = The element of $\mathbf{A}$ existing in row i and column j.

The dimensions of a matrix are

$$(number\ of\ rows) \times (number\ of\ columns) = r \times c.$$

A **vector** is a matrix with dimensions $(r \times 1)$.

Example 7

Matrix operations

- Matrix addition and subtraction

DEFINITION 3.2 Let $\mathbf{A}$ and $\mathbf{B}$ be matrices that have the same dimensions ($r \times c$). Then

$$\mathbf{A}+\mathbf{B} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1c} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{i1} & b_{i2} & \cdots & b_{ij} & \cdots & b_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{r1} & b_{r2} & \cdots & b_{rj} & \cdots & b_{rc} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1j}+b_{1j} & \cdots & a_{1c}+b_{1c} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2j}+b_{2j} & \cdots & a_{2c}+b_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1}+b_{i1} & a_{i2}+b_{i2} & \cdots & a_{ij}+b_{ij} & \cdots & a_{ij}+b_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1}+b_{r1} & a_{r2}+b_{r2} & \cdots & a_{rj}+b_{rj} & \cdots & a_{rc}+b_{rc} \end{pmatrix}.$$

Similarly,

$$\mathbf{A}-\mathbf{B} = \begin{pmatrix} a_{11}-b_{11} & a_{12}-b_{12} & \cdots & a_{1j}-b_{1j} & \cdots & a_{1c}-b_{1c} \\ a_{21}-b_{21} & a_{22}-b_{22} & \cdots & a_{2j}-b_{2j} & \cdots & a_{2c}-b_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1}-b_{i1} & a_{i2}-b_{i2} & \cdots & a_{ij}-b_{ij} & \cdots & a_{ij}-b_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1}-b_{r1} & a_{r2}-b_{r2} & \cdots & a_{rj}-b_{rj} & \cdots & a_{rc}-b_{rc} \end{pmatrix}.$$

- Matrix multiplication

Matrix multiplication is only defined if the *inside* dimensions match. If $\mathbf{A}$ has dimensions $r \times c$ and $\mathbf{B}$ has dimensions $c \times k$, then the product $\mathbf{A} \cdot \mathbf{B}$ can be computed. The resulting matrix $\boldsymbol{D} = \mathbf{A} \cdot \mathbf{B}$ has dimensions $r \times k$.

DEFINITION 3.3 Let $\mathbf{A}$ be a matrix with dimensions $r \times c$ and $\mathbf{B}$ be a matrix with dimensions $c \times k$. Then the product of $\mathbf{A}$ and $\mathbf{B}$ is defined by

$$\mathbf{A} \cdot \mathbf{B} = \boldsymbol{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1j} & \cdots & d_{1k} \\ d_{21} & d_{22} & \cdots & d_{2j} & \cdots & d_{2k} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{i1} & d_{i2} & \cdots & d_{ij} & \cdots & d_{ik} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{r1} & d_{r2} & \cdots & d_{rj} & \cdots & d_{rk} \end{pmatrix},$$

where the $\{ij\}^{th}$ element of $\boldsymbol{D}$ is computed by

$$d_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{rj}.$$

- Scalar multiplication

DEFINITION 3.4 Let $\mathbf{A}$ be a matrix with dimensions $r \times c$ and let q be a real number. Then

$$q\mathbf{A} = \begin{pmatrix} qa_{11} & qa_{12} & \cdots & ra_{1j} & \cdots & qa_{1c} \\ qa_{21} & qa_{22} & \cdots & qa_{2j} & \cdots & qa_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ qa_{i1} & qa_{i2} & \cdots & qa_{ij} & \cdots & qa_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ qa_{r1} & qa_{r2} & \cdots & qa_{rj} & \cdots & qa_{rc} \end{pmatrix}.$$

- Transpose of a matrix

DEFINITION 3.5 Let $\mathbf{A}$ be a matrix with dimensions $r \times c$. The **transpose** of matrix $\mathbf{A}$ denoted $\mathbf{A}^T$ (or $\mathbf{A}'$) is defined by

$$\mathbf{A}^T = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{i1} & \cdots & a_{r1} \\ a_{12} & a_{22} & \cdots & a_{i2} & \cdots & a_{r2} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{1j} & a_{2j} & \cdots & a_{ji} & \cdots & a_{rj} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{1c} & a_{2c} & \cdots & a_{ic} & \cdots & a_{cr} \end{pmatrix}.$$

Special matrices

- The identity matrix

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

Note: For any matrix $\mathbf{A}$,

$$\mathbf{I} \cdot \mathbf{A} = \mathbf{A} \quad \text{and} \quad \mathbf{A} \cdot \mathbf{I} = \mathbf{A}$$

- The one-vector and one-matrix.

$$\mathbf{1}_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \mathbf{J}_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 1 & \cdots & 1 & 1 \end{pmatrix}$$

Inverse of a matrix

Consider solving the equation $3x = 6$.

DEFINITION 3.6 For a square matrix $\mathbf{A}$, the inverse denoted $\mathbf{A}^{-1}$, is a matrix that satisfies

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{I}.$$

Non-square matrices do not have inverses. Also, not all square matrices have inverses. A square matrix that does not have an inverse is called **singular**.

Inverse of a 2×2 matrix

Solving systems of equations

Example 8

Consider the system of equations

$$\begin{cases} 4x + 2y - z = 12 \\ 3x + y + z = 4 \\ x - y - 2z = 2 \end{cases}$$

R code:

```
A <- matrix(c(4,3,1,2,1,-1,-1,1,-2),nrow=3,ncol=3)
b <- c(12,4,2)
solve(A,b)
```

```
[1]  1  3 -2
```

3.2 Random Vectors and Matrices

DEFINITION 3.7 A **random vector** or a **random matrix** is a rectangular array of random variables.

- Let $\mathbf{Y} = \begin{pmatrix} Y_1 & Y_2 & \cdots & Y_n \end{pmatrix}^T$ be a random vector. Then the expected value and covariance matrix of $\mathbf{Y}$ are respectively

$$E(\mathbf{Y}) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix}, \quad Var(\mathbf{Y}) = \begin{pmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \cdots & Var(Y_n) \end{pmatrix}.$$

The covariance matrix is also defined by

$$Var(\mathbf{Y}) = E\{[\mathbf{Y} - E(\mathbf{Y})][\mathbf{Y} - E(\mathbf{Y})]^T\}.$$

Note:

3.3 Matrix Form of the Multiple Linear Regression Model

Before introducing the regression model, we must first define the design matrix.

DEFINITION 3.8 Consider a data set consisting of $p-1$ covariates $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{p-1}$. Then, the **design matrix** is defined by

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_n & \mathbf{X}_1 & \mathbf{X}_2 & \ldots, & \mathbf{X}_{p-1} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{pmatrix}. \tag{3.3}$$

Matrix multiple linear regression model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3.4}$$

where $\mathbf{Y}$ is the response vector, $\mathbf{X}$ is the design matrix from Equation (3.3), $\boldsymbol{\beta}$ is the parameter vector

$$\boldsymbol{\beta} = \begin{pmatrix}\beta_0 & \beta_1 & \cdots & \beta_{p-1}\end{pmatrix}^T$$

and $\boldsymbol{\epsilon}$ is a vector of independent normal random variables with expectation $E[\boldsymbol{\epsilon}] = 0$ and covariance matrix

$$Var[\boldsymbol{\epsilon}] = \sigma^2 I_n = \begin{pmatrix}\sigma^2 & 0 & \cdots & 0\\ 0 & \sigma^2 & \cdots & \vdots\\ \vdots & \vdots & \ddots & 0\\ 0 & \cdots & 0 & \sigma^2\end{pmatrix}.$$

3.4 Estimation of the Multiple Linear Regression Model

Recall for simple linear regression, the least squares estimators are derived by minimizing

$$Q(b_0, b_1) = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_i))^2,$$

with respect to b_0 and b_1. We need to come up with an analogous criterion using the matrix formulation of the multiple regression model. First, define

$$Q(b_0, b_1, b_2, \ldots, b_{p-1}) = \sum_{i=1}^{n}(y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_{p-1}x_{i,p-1}))^2, \tag{3.5}$$

and

$$\mathbf{b} = \begin{pmatrix}b_0 & b_1 & b_2 & \ldots & b_{p-1}\end{pmatrix}^T.$$

Then Q can be expressed

$$Q(b_0, b_1, b_2, \ldots, b_{p-1}) = Q(\mathbf{b}) = (\mathbf{Y} - \mathbf{X}\mathbf{b})^T(\mathbf{Y} - \mathbf{X}\mathbf{b}). \tag{3.6}$$

To derive the least squares estimators of the multiple linear regression model, we will minimize Q with respect to the vector $\mathbf{b}$. This result is summarized in the following proposition.

PROPOSITION 3.1 *Let Q be defined in Equation (3.5) or (3.6). Then Q is minimized when*

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

Denote the minimum by $\widehat{\boldsymbol{\beta}}$. Hence, the least squares estimator of the multiple linear regression model is computed by

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{3.7}$$

Further, the minimum value of Q is

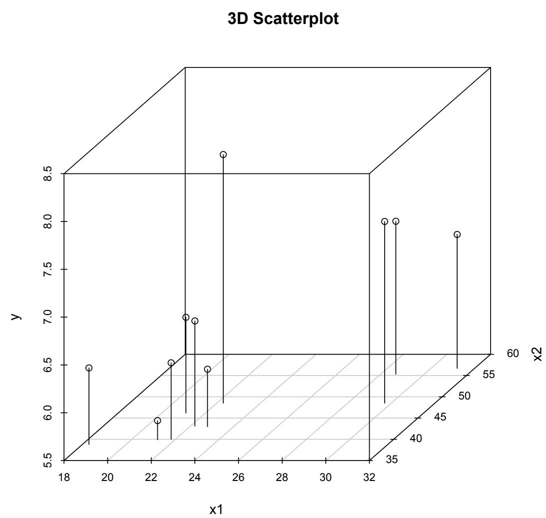
$$SSE = Q(\widehat{\boldsymbol{\beta}}) = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})^T(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}). \tag{3.8}$$

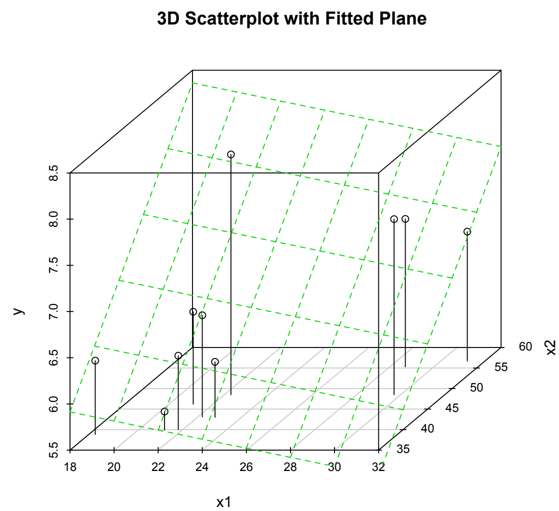
Derivation of the least squares Equation (3.7) follows:

Example 9

The State of Vermont is divided into 10 Health Planning Districts – they correspond roughly to counties. The following data represent the percentage of live births of babies weighing under 2500 grams (y), the fertility rate for females younger than 18 or older than 34 years of age (x_1), total high-risk fertility rate for females younger than 17 or older than 35 years of age (x_2), percentage of mothers with fewer than 12 years of education (x_3), percentage of births to unmarried mothers (x_4), and percentage of mothers not seeking medical care until the third trimester (x_5).

	y	x_1	x_2	x_3	x_4	x_5	$(y-\bar{y})^2$	$(y-\hat{y})^2$
	6.1	43.0	22.8	23.8	9.2	6		
	7.1	55.3	28.7	24.8	12.0	10		
	7.4	48.5	27.7	23.9	10.4	5		
	6.3	38.8	18.3	16.6	9.8	4		
	6.5	46.2	21.1	19.6	9.8	5		
	5.7	39.9	21.2	21.4	7.7	6		
	6.6	43.1	22.2	20.7	10.9	7		
	8.1	48.5	22.3	21.8	5.5	5		
	6.3	40.0	21.8	20.6	11.6	7		
	6.9	56.7	31.2	25.2	11.6	9		
Total								





3D scatter plots using response variable y and covariates x_1, x_2 are displayed above. The statistical model for this application is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

with

$$\epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

The figure on the right shows the estimated model (*estimated plane*) with the raw data. To find the estimated model, we will use the least squares Equation (3.7).

The estimated model is: $\hat{y} = 3.369 - 0.049x_1 + 0.098x_2$

The `R` code follows:

```
y <- c(6.1,7.1,7.4,6.3,6.5,5.7,6.6,8.1,6.3,6.9)
x1 <- c(22.8,28.7,29.7,18.3,21.1,21.2,22.2,22.3,21.8,31.2)
x2 <- c(43,55.3,48.5,38.8,46.2,39.9,43.1,48.5,40,56.7)
n <- length(y)
```

Matrix approach:

```
X <- cbind(rep(1,n),x1,x2)
solve(t(X)%*%X)%*%t(X)%*%y
```

Approach using `lm( )`:

```
lm(y~x1+x2)
```

Data set was loaded from a tex or csv file: (class data.frame)

```
lm(y~x1+x2,data=name)
```

Now assume that we include all covariates x_1, x_2, x_3, x_4 and x_5. Then the statistical model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon$$

with

$$\epsilon \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

Again we use the least squares Equation 3.7 to find the estimated model.

Matrix approach:

```
X <- cbind(rep(1,n),x1,x2,x3,x4,x5)
solve(t(X)%*%X)%*%t(X)%*%y
```

Approach using `lm( )`:

```
lm(y~x1+x2+x3+x4+x5)
```

The estimated *full* model is:

$$\hat{y} = 5.3590 + 0.0840x_1 + 0.0873x_2 - 0.1230x_3 - 0.1510x_4 - 0.0801x_5$$