Regression 5

Matrix Approaches to Simple Linear Regression Analysis

1 Basics on matrix algebra

Read Section 5.1 $\sim$ 5.7 and refer to a basic linear algebra textbook.

2 Basics on Vector

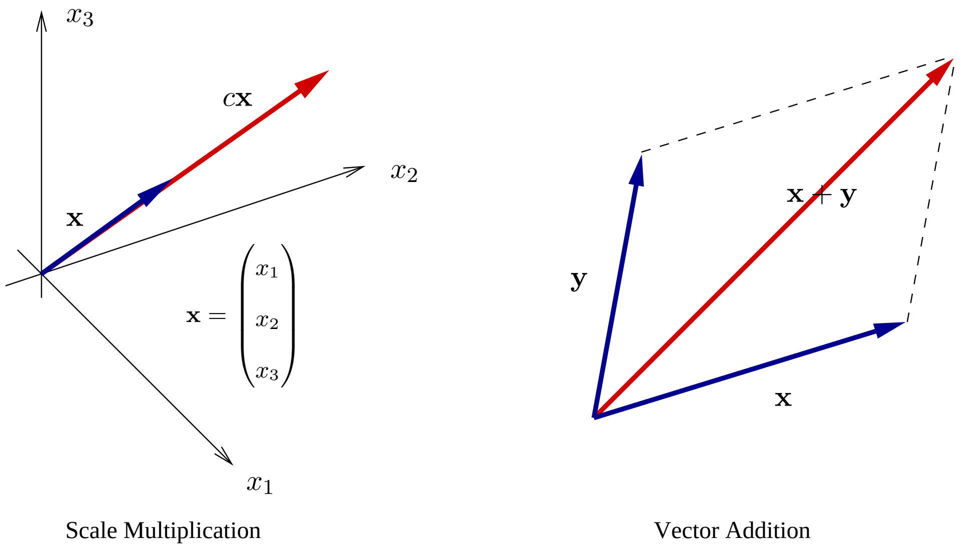


Figure 5.1: Vector Operations

- The scalar multiplication of a vector by a scalar, $c\mathbf{x}$, is defined as

$$c\mathbf{x} = c\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} cx_1 \\ cx_2 \\ cx_3 \end{pmatrix}.$$

- The vector addition, $\mathbf{x}+\mathbf{y}$, is defined as

$$\mathbf{x}+\mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ x_3+y_3 \end{pmatrix}.$$

- The vector subtraction, $\mathbf{x}-\mathbf{y}$, is defined as

$$\mathbf{x}+(-1)\mathbf{y}.$$

- The length (norm, magnitude) of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ is defined as

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

  Note that $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$. A vector $\mathbf{x}$ of length 1 is called a unit vector. Any non-zero vector can be written as a scale multiplication of a *unit* vector by its length, *i.e.*, $\mathbf{x} = c\mathbf{u}$, where $c = \|\mathbf{x}\|$ and $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$.

- The inner (dot *or* scalar) product of two vectors, $\mathbf{x} = (x_1, \ldots, x_n)'$ and $\mathbf{y} = (y_1, \ldots, y_n)'$ is defined as

$$\mathbf{x}'\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \circ \mathbf{y} = x_1y_1 + \cdots + x_ny_n = \sum_{i=1}^{n} x_iy_i.$$

  The angle θ between two non-zero vectors $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\cos\theta = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} = \mathbf{u}'\mathbf{v},$$

  where $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$ and $\mathbf{v} = \mathbf{y}/\|\mathbf{y}\|$.

  Note that $\mathbf{x}'\mathbf{y} = 0$ implies that these two vectors are orthogonal.

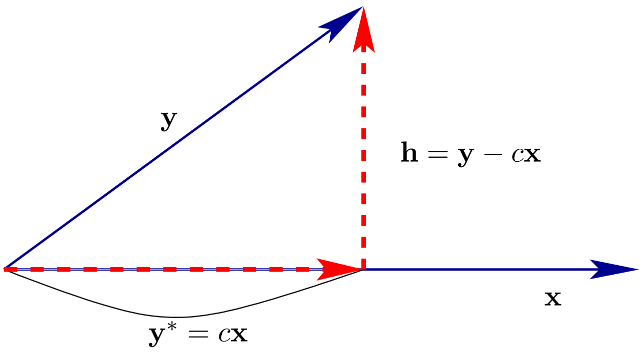


Figure 5.2: Orthogonal projection of $\mathbf{y}$ on $\mathbf{x}$.

Example 5.1. The sample correlation of a set of observations $\{(x_i, y_i) : i = 1, \ldots, n\}$ is given by the formula

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

The above formula can be represented by the angle between two vectors

$$\cos\theta = \frac{\mathbf{x}^{*\prime}\mathbf{y}^*}{\|\mathbf{x}^*\| \cdot \|\mathbf{y}^*\|},$$

where $\mathbf{x}^* = (x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_n - \bar{x})'$ and $\mathbf{y}^* = (y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y})'$. $\|$

- Orthogonal projection of a vector onto a line as seen in Figure 5.2.

  The vector denoted by $\mathbf{y}^*$ is the orthogonal projection of $\mathbf{y}$ on $\mathbf{x}$. To find this vector, we need to find a constant c. Since $\mathbf{x}$ and $\mathbf{h}$ are orthogonal, we have $\mathbf{x}'\mathbf{h} = \mathbf{x}'(\mathbf{y} - c\mathbf{x}) = 0$ resulting in $\mathbf{x}'\mathbf{y} - c\mathbf{x}'\mathbf{x} = 0$. It follows that

  $$c = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}.$$

  The orthogonal projection of $\mathbf{y}$ onto $\mathbf{x}$, denoted by $\mathbf{y}^*$, is given by

  $$\boxed{\mathbf{y}^* = c\mathbf{x} = \left(\frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}\right)\mathbf{x} = \mathbf{x}\left(\frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}\right) = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}.}$$

Example 5.2. The sample mean can be represented by the orthogonal projection as seen in Figure 5.2. Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ and $\mathbf{x} = (1, 1, \ldots, 1)'$. Then the orthogonal projection

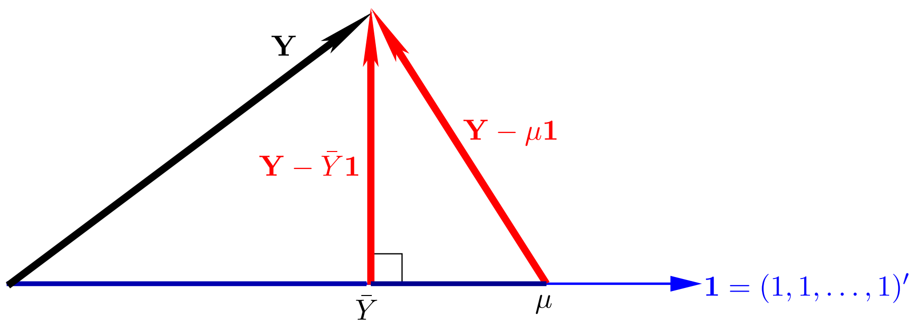


Figure 5.3: Orthogonal projection of $\mathbf{Y}$ on $\mathbf{1}=(1,1,\ldots,1)'$.

of $\mathbf{y}$ onto $\mathbf{1}=(1,1,\ldots,1)'$ is given by

$$\mathbf{y}^* = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\mathbf{y} = \mathbf{1}(n)^{-1}\mathbf{1}'\mathbf{y} = \tfrac{1}{n}\mathbf{J}\mathbf{y} = \bar{y}\mathbf{1},$$

where $\mathbf{J}=\mathbf{1}\mathbf{1}'$ (all-ones matrix). $\|$

Lemma 5.1. *Let Z_i be independent and normally distributed random variables with $N(0,1)$ for $i=1,2,\ldots,\nu$ and $\sum_{i=1}^{\nu} Z_i^2 = V_1+V_2+\cdots+V_s$, where V_j has ν_j degrees of freedom for $j=1,2,\ldots,s$ and $\nu_j>0$. Then V_j are independent with chi-square random variables each with ν_j degrees of freedom* if and only if $\nu=\nu_1+\nu_2+\cdots+\nu_s$.

Proof. Originally from Cochran (1934). Also, see Theorem 1 in Appendix VI of Scheffé (1959) and Chapter 15 of Kendall and Stuart (1979). □

Example 5.3. The chi-square decomposition is also represented by the orthogonal projection as seen in Figure 5.3. Let $\mathbf{Y}=(Y_1,Y_2,\ldots,Y_n)'\sim \mathrm{MVN}(\mu\mathbf{1},\sigma^2\mathbf{I})$. Then the orthogonal projection of the random vector $\mathbf{Y}$ onto $\mathbf{1}=(1,1,\ldots,1)'$ is given by $\bar{Y}\mathbf{1}$. It is easily seen that

$$\|\mathbf{Y}\|^2 = \|\mathbf{Y}-\bar{Y}\mathbf{1}\|^2 + \|\bar{Y}\mathbf{1}\|^2$$

and

$$\|\mathbf{Y}-\mu\mathbf{1}\|^2 = \|\mathbf{Y}-\bar{Y}\mathbf{1}\|^2 + \|(\bar{Y}-\mu)\mathbf{1}\|^2,$$

which are respectively equivalent to

$$\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 + n\bar{Y}^2.$$

and

$$\sum_{i=1}^n (Y_i - \mu)^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2 + n(\bar{Y} - \mu)^2.$$

Thus, we have the following chi-square decomposition.

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - \mu)^2}_{\sim\ \chi^2_{\mathrm{df}=n}} = \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\sim\ \chi^2_{\mathrm{df}=n-1}} + \underbrace{\frac{1}{\sigma^2} n(\bar{Y} - \mu)^2}_{\sim\ \chi^2_{\mathrm{df}=1}}.$$

We have the following F-test statistic

$$F = \frac{\frac{1}{\sigma^2} n(\bar{Y} - \mu)^2/1}{\frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - \bar{Y})^2/(n-1)} = \frac{n(\bar{Y} - \mu)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2/(n-1)} \sim F(1, n-1).$$

It is easily seen that $F = T^2$ and

$$T = \frac{\bar{Y} - \mu}{\sqrt{\mathrm{MSE}/n}} \sim t_{\mathrm{df}=n-1},$$

where $\mathrm{MSE} = \sum_{i=1}^n (Y_i - \bar{Y})^2/(n-1)$. ||

- Orthogonal projection of a vector onto a space.

  Is it possible to project a vector onto a space spanned by vectors? See Theorem 4.4.1 of Graybill (1983).

  Let $\mathbf{y}^*$ denote the orthogonal projection of $\mathbf{y}$ onto the space spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$. Then the orthogonal projection is given by

  $$\boxed{\mathbf{y}^* = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},}$$

  where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2)$.

  Note that $\|\mathbf{y} - \mathbf{y}^*\|$ is the shortest distance from the point given by $\mathbf{y}$ to the space.

3 Random vectors and random matrices

A *random vector* is a vector whose elements are random variables. Similarly, a *random matrix* is a matrix whose elements are random variables.

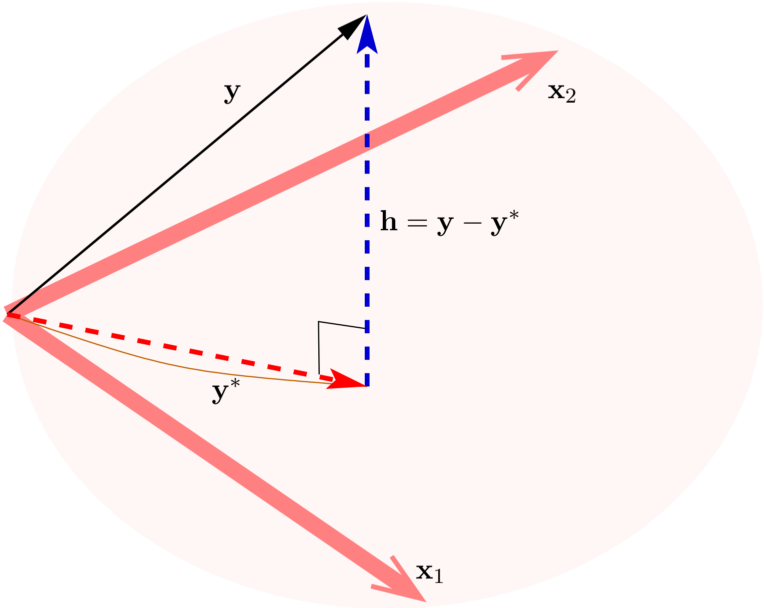


Figure 5.4: Orthogonal projection of $\mathbf{y}$ on $\mathcal{R}(\mathbf{x_1}, \mathbf{x_2})$.

3.1 Expectation of Random Vector or Matrix

The expected value of a random matrix (or vector) is the matrix consisting of the expected values of each of its elements. Specifically, let $\mathbf{Y}$ be a $n \times 1$ random vector. Then the random vector $\mathbf{Y}$ and the expected value of $\mathbf{Y}$ are

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \text{and} \quad \underset{n\times 1}{E(\mathbf{Y})} = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix}.$$

For a random matrix $\mathbf{Y}$ with dimension $n \times p$, the random matrix and the expected value of $\mathbf{Y}$ are

$$\underset{n\times p}{\mathbf{Y}} = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1p} \\ Y_{21} & Y_{22} & \cdots & Y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{np} \end{bmatrix} \quad \text{and} \quad \underset{n\times p}{E(\mathbf{Y})} = \begin{bmatrix} E(Y_{11}) & E(Y_{12}) & \cdots & E(Y_{1p}) \\ E(Y_{21}) & E(Y_{22}) & \cdots & E(Y_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(Y_{n1}) & E(Y_{n2}) & \cdots & E(Y_{np}) \end{bmatrix}.$$

Example 5.4. Suppose that ϵ_i are error terms with $E[\epsilon_i] = 0$. Let $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]'$. Then we have

$$\underset{n\times 1}{E[\boldsymbol{\epsilon}]} = \underset{n\times 1}{\mathbf{0}}.$$

$\|$

3.2 Variance-Covariance matrix of random vector

The variance-covariance matrix of an $n \times 1$ random vector $\mathbf{Y}$ is defined as

$$\underbrace{\mathrm{Cov}(\underset{n\times 1}{\mathbf{Y}})}_{n\times n} \equiv E\Big[\big(\mathbf{Y} - E(\mathbf{Y})\big)\big(\mathbf{Y} - E(\mathbf{Y})\big)'\Big] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix},$$

where $\sigma_{ij} = \mathrm{Cov}(Y_i, Y_j) = E\big[(Y_i - E(Y_i))(Y_j - E(Y_j))\big]$.

Note that $\sigma_{ii} = \mathrm{Cov}(Y_i, Y_i) = \mathrm{Var}(Y_i)$ and $\sigma_{ij} = \sigma_{ji}$.

Example 5.5. Suppose that ϵ_i are *iid* with $\mathrm{Var}(\epsilon_i) = \sigma^2$. Let $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]'$. Then $\mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$. $\|$

3.3 Some basic theorems

Theorem 5.2. *Let $\mathbf{A}$ and $\mathbf{B}$ be constant matrices and $\mathbf{X}$ or $\mathbf{Y}$ be a random vector or matrix. Then we have:*

1. $E(\mathbf{A}) = \mathbf{A}$.
2. $E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y})$.
3. $E(\mathbf{AYB}) = \mathbf{A}E(\mathbf{Y})\mathbf{B}$.
4. $\mathrm{Cov}(\mathbf{Y}) = E(\mathbf{YY}') - \boldsymbol{\mu}_Y\boldsymbol{\mu}_Y'$, *where* $\boldsymbol{\mu}_Y = E(\mathbf{Y})$.
5. $\mathrm{Cov}(\mathbf{AY}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{Y})\mathbf{A}'$.

4 Simple linear regression model in matrix terms

Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \tag{5.1}$$

where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ and $i = 1, \ldots, n$. This implies

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_1 + \epsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_2 + \epsilon_2 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_n + \epsilon_n
\end{aligned}$$

Let us define $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{\beta}$ as follows:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \underset{n\times 1}{\boldsymbol{\epsilon}} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \text{ and } \underset{2\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Using these vectors, we can rewrite (5.1) as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

1. $\mathbf{Y}$ is a random vector which concerns the response variables.
2. $\mathbf{X}$ is a matrix of data which concerns the predictor variables.

   $\mathbf{X}$ is assumed known (fixed).
3. $\boldsymbol{\beta}$ is the parameter vector.
4. $\boldsymbol{\epsilon}$ is a random vector such that $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$.

   Note that $\text{Cov}(\mathbf{Y}) = \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}$.

5 Least-squares estimation of simple linear regression parameters

From the normal equations of Chapter 1, we have

$$\frac{\partial Q_2}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i) = 0.$$
$$\frac{\partial Q_2}{\partial \beta_1} = -2\sum_{i=1}^{n}X_i(Y_i - \beta_0 - \beta_1 X_i) = 0.$$

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the solution of the above equations and it follows that

$$\begin{aligned} n\hat{\beta}_0 + (\sum X_i)\hat{\beta}_1 &= \sum Y_i \\ (\sum X_i)\hat{\beta}_0 + (\sum X_i^2)\hat{\beta}_1 &= \sum X_i Y_i, \end{aligned}$$

and in matrix notation

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}.$$

It follows that

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y},$$

where $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0, \hat{\beta}_1]'$,

$$\underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \text{and} \quad \underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}.$$

Hence we have

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

6 Fitted values and residuals

The fitted values $\hat{Y}_i$ are

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \qquad i = 1, \ldots, n.$$

In matrix notation, we have

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}},$$

where $\hat{\mathbf{Y}} = [\hat{Y}_1, \ldots, \hat{Y}_n]'$ and $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0, \hat{\beta}_1]'$. Since $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, we have

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y},$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. The $n \times n$ matrix $\mathbf{H}$ is called the hat matrix. The hat matrix $\mathbf{H}$ has the following properties:

(i) Idempotent: $\mathbf{HH} = \mathbf{H}$.

(ii) Symmetric: $\mathbf{H}' = \mathbf{H}$.

Let the vector of the residuals $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ be denoted by

$$\underset{n\times 1}{\hat{\boldsymbol{\epsilon}}} = \begin{bmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{bmatrix}.$$

Then we have

$$\underset{n\times 1}{\hat{\boldsymbol{\epsilon}}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}.$$

Note that $(\mathbf{I} - \mathbf{H})$ is also (i) symmetric and (ii) idempotent. The covariance matrix of the vector of residuals $\hat{\boldsymbol{\epsilon}}$ is

$$\operatorname{Cov}(\hat{\boldsymbol{\epsilon}}) = \operatorname{Cov}((\mathbf{I} - \mathbf{H})\mathbf{Y}) = (\mathbf{I} - \mathbf{H})\operatorname{Cov}(\mathbf{Y})(\mathbf{I} - \mathbf{H})' = \sigma^2(\mathbf{I} - \mathbf{H}),$$

and is estimated by

$$\widehat{\operatorname{Cov}}(\hat{\boldsymbol{\epsilon}}) = \mathrm{MSE}\,(\mathbf{I} - \mathbf{H}).$$

Remark 5.1.

1. A square matrix $\mathbf{P}$ is a projection matrix if and only if $\mathbf{P}^2 = \mathbf{P}$. Thus, the hat matrix $\mathbf{H}$ above is a projection matrix.

2. In addition, a projection matrix $\mathbf{P}$ is an orthogonal projector if $\mathbf{P}$ is symmetric. This is easily shown as follows. Let $\mathbf{x}$ and $\mathbf{y}$ be arbitrary n-dimensional vectors. Suppose that $\mathbf{P}$ is an orthogonal projector. Then we have

$$\langle \mathbf{Px}, \mathbf{y} - \mathbf{Py} \rangle = \langle \mathbf{Py}, \mathbf{x} - \mathbf{Px} \rangle = 0,$$

which results in

$$\langle \mathbf{Px}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{Py} \rangle.$$

Thus, we have $\mathbf{x}'\mathbf{P}'\mathbf{y} = \mathbf{x}'\mathbf{Py}$. Since $\mathbf{x}'\mathbf{P}'\mathbf{y} = \mathbf{x}'\mathbf{Py}$ holds for any $\mathbf{x}$ and $\mathbf{y}$, we have $\mathbf{P}' = \mathbf{P}$, which implies that $\mathbf{P}$ should be symmetric. Since $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is symmetric, it is an orthogonal projector or an orthogonal projection matrix.

3. Note that an idempotent but asymmetric matrix is an oblique projection matrix. In the weighted least squares (WLS) regression, the hat matrix is given by $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}$, which is idempotent but not symmetric. Thus, it is an oblique projection matrix. This topic will be covered in Chapter 11.

△

Example 5.6. Matrix example. A company produces refrigeration equipment and its replacement parts. In the past, one of the replacement parts has been produced periodically in different size lots. The company is interested in the optimum lot size. The data in the first column are different lot sizes and those in the second column are their corresponding work hours required to produce the lot (Kutner et al., 2005).

Minitab

Read Data

```
# Editor -> Enable Commands
MTB > read c1 c2;
SUBC>      file "S:\LM\CH01TA01.txt" .
Entering data from file: S:\LM\CH01TA01.TXT
25 rows read.
```

Regression

```
MTB > regr c2 1 c1;
SUBC>     resid c3;
SUBC>     fits  c4.

Regression Analysis: C2 versus C1

The regression equation is
C2 = 62.4 + 3.57 C1

```

```
Predictor      Coef   SE Coef       T      P
Constant      62.37    26.18     2.38  0.026
C1            3.5702   0.3470   10.29  0.000

S = 48.8233   R-Sq = 82.2%   R-Sq(adj) = 81.4%

Analysis of Variance
Source           DF      SS      MS       F      P
Regression        1  252378  252378  105.88  0.000
Residual Error   23   54825    2384
Total            24  307203
```

Matrix calculations

```
MTB > let k1 = count(c1)

# Make a one vector in c11
MTB > set c11
DATA>  k1(1)
DATA> end

MTB > copy  c11 c1  m1        ## m1  = X  matrix
MTB > transpose m1  m11       ## m11 = X' matrix
MTB > multiply  m11 m1  m2    ## m2  = X'X
MTB > invert    m2  m3        ## m3  =(X'X)^(-1)
MTB > multiply  m11 c2  c5    ## c5  = X'Y
MTB > multiply  m3  c5  c6    ## c6  = (X'X)^(-1) X'Y (beta hat)
MTB > multiply  m1  m3  m12   ## m12 = X(X'X)^(-1)
MTB > multiply  m12 m11 m4    ## m4  = X(X'X)^(-1)X' (H matrix)
MTB > multiply  m4  c2  c44   ## c44 = Y hat

# Read or Print a matrix,
MTB > print m3
Data Display
 Matrix M3
 0.287475   -0.0035354
-0.003535    0.0000505

## The following is (b0, b1).
MTB > print c6

Data Display
C6
   62.3659    3.5702

MTB > ## C44 should be the same as the C4
MTB > print c44 c4

Data Display
Row      C44       C4
  1  347.982  347.982
  2  169.472  169.472
  3  240.876  240.876
  4  383.684  383.684
  5  312.280  312.280
  6  276.578  276.578
  7  490.790  490.790
  8  347.982  347.982
  9  419.386  419.386
   ...................

MTB > let c33 = c2 - c44  ## c33 = e hat

## C33 should be the same as the C3
MTB > print c33 c3

Data Display
Row      C33       C3
  1   51.018   51.018
  2  -48.472  -48.472
  3  -19.876  -19.876
```

```
  4   -7.684   -7.684
  5   48.720   48.720
  6  -52.578  -52.578
  7   55.210   55.210
  8    4.018    4.018
  9  -66.386  -66.386
  .....................

MTB > let k11 = sum(c11*c33)  ## 1' * e-hat
MTB > let k12 = sum(c1 *c33)  ## x1 * e-hat

## The followings should be zero.
MTB > print k11 k12

Data Display
K11     0.000000000
K12     0.000000000
```

R

Read Data

```
> mydata =
    read.table("https://raw.githubusercontent.com/AppliedStat/LM/master/CH01TA01.txt")
>
> ## On purpose, I used the same symbols as the minitab variables.
>
> c1 = mydata[,1]
> c2 = mydata[,2]
```

Regression

```
> LM = lm (c2 ~ c1)
> c3 = resid(LM)
> c4 = fitted(LM)
>
> summary(LM)

Call:
lm(formula = c2 ~ c1)

Residuals:
    Min       1Q  Median       3Q      Max
-83.876 -34.088  -5.982  38.826 103.528

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   62.366     26.177   2.382   0.0259 *
c1             3.570      0.347  10.290 4.45e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 48.82 on 23 degrees of freedom
Multiple R-Squared: 0.8215, Adjusted R-squared: 0.8138
F-statistic: 105.9 on 1 and 23 DF,  p-value: 4.449e-10
```

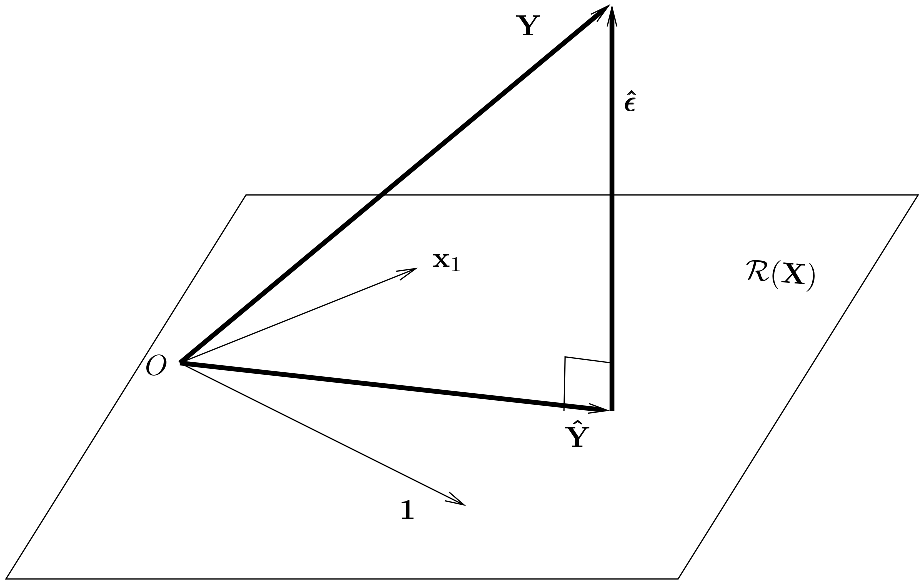
Matrix calculations

```
> k1  = length(c1)
> c11 = rep(1,k1)

> m1  = cbind(c11, c1) ## m1 = X matrix
> m11 = t(m1)          ## m11 = X' matrix
> m2  = m11 %*% m1     ## m2  = X'X  Note: m2 = t(m1) %*% m1
> m3  = solve(m2)      ## (X'X)^(-1)

> ## Check m2 %*% m3. It should give the 2x2 identity matrix.
> c5  = m11 %*% c2     ## c5 = X'Y
```

```
> c6  = m3  %*% c5     ## c6 = (X'X)^(-1) X'Y (beta hat)
>
> m12 = m1  %*% m3     ## m12 = X (X'X)^(-1)
> m4  = m12 %*% m11    ## m4  = X (X'X)^(-1) X' (H matrix)

> ## Check HHHH..... eg.: m4 %*% m4 %*% m4 - m4,
> m4 %*% m4 %*% m4 - m4
               [,1]          [,2]          [,3]          [,4]          [,5]
 [1,] -8.326673e-17 -4.163336e-17 -4.857226e-17 -7.632783e-17 -5.551115e-17
 [2,] -6.245005e-17 -1.804112e-16 -1.526557e-16 -3.095397e-17 -9.020562e-17
 [3,] -5.551115e-17 -1.110223e-16 -9.020562e-17 -4.857226e-17 -6.938894e-17
 [4,] -7.632783e-17 -1.149254e-17 -4.163336e-17 -7.632783e-17 -5.551115e-17
 [5,] -5.551115e-17 -6.245005e-17 -6.938894e-17 -6.245005e-17 -5.551115e-17
           ..................

> c44 = m4 %*% c2      ## X (X'X)^(-1) X' Y = Y hat
> ## Note: In R, we can do the above in a single line.
> ##       Of course, we can use any alpha-numeric variable names.
> X=m1; Y=c2; c4tmp = X %*% solve( t(X)%*%X ) %*% t(X) %*% Y

> ## Compare beta's
> coefficients(LM)
(Intercept)          c1
  62.365859    3.570202

> c6
          [,1]
c11 62.365859
c1   3.570202

> ## c44 and c4tmp should be the same as the c4
> cbind(c44, c4tmp, c4)
                           c4
1  347.9820 347.9820 347.9820
2  169.4719 169.4719 169.4719
3  240.8760 240.8760 240.8760
4  383.6840 383.6840 383.6840
5  312.2800 312.2800 312.2800
6  276.5780 276.5780 276.5780
7  490.7901 490.7901 490.7901
8  347.9820 347.9820 347.9820
9  419.3861 419.3861 419.3861
   .....................

> c33 = c2 - c4  ## c33 = e hat
> cbind(c33,c3)
          c33          c3
1   51.0179798  51.0179798
2  -48.4719192 -48.4719192
3  -19.8759596 -19.8759596
4   -7.6840404  -7.6840404
5   48.7200000  48.7200000
6  -52.5779798 -52.5779798
7   55.2098990  55.2098990
8    4.0179798   4.0179798
9  -66.3860606 -66.3860606
   ...................

> k11 = sum(c11*c33)  ## 1' * e-hat
> k12 = sum(c1 *c33)  ## x1 * e-hat
>
> ## The followings should be zero.
> c(k11, k12)
[1] 2.842171e-13 1.818989e-11
```

7 The geometry of the least squares procedures

The geometry of the least squares procedures



If $\mathbf{X}$ is an $n \times 2$ matrix

$$\underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{x}_1 \end{bmatrix},$$

and $\mathbf{1}$ and $\mathbf{x}_1$ are the column vectors of $\mathbf{X}$, which are $\mathbf{1} = [1, \ldots, 1]'$ and $\mathbf{x}_1 = [X_1, \ldots, X_n]'$. The space spanned by the column vectors of $\mathbf{X}$ is the set of all vectors that can be written as

$$c_0\mathbf{1} + c_1\mathbf{x}_1,$$

where c_0 and c_1 are any real numbers. In this case, this space is a plane passing through $\mathbf{1}$ and $\mathbf{x}_1$, and the origin. The space spanned by the column vectors of $\mathbf{X}$ is sometimes called the range space of $\mathbf{X}$ and is denoted by $\mathcal{R}(\mathbf{X})$. The basic idea of regression is to find the point within $\mathcal{R}(\mathbf{X})$ that is closest to $\mathbf{Y}$, that is, to choose a value of $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ so that

$$\mathbf{X}\boldsymbol{\beta} = \beta_0\mathbf{1} + \beta_1\mathbf{x}_1$$

is closest to $\mathbf{Y}$. We will choose this value $\hat{\boldsymbol{\beta}}$ so that $\mathbf{X}\hat{\boldsymbol{\beta}}$ is the projection of $\mathbf{Y}$ onto $\mathcal{R}(\mathbf{X})$. This will result in $\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ having minimum Euclidean length, or $||\hat{\boldsymbol{\epsilon}}||^2 = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}$ being minimized. One can show that the projection matrix onto $\mathcal{R}(\mathbf{X})$ is given by

$$\mathbf{P}_{\mathcal{R}(\mathbf{X})} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

The projection *or* shadow of $\mathbf{Y}$ onto $\mathcal{R}(\mathbf{X})$ is

$$\mathbf{P}_{\mathcal{R}(\mathbf{X})}\mathbf{Y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}.$$

Note that the projection of $\mathbf{Y}$ onto $\mathcal{R}(\mathbf{X})$,

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

is a generalization of vector projection while the projection of $\mathbf{y}$ onto $\mathbf{x}$ was

$$\mathbf{y}^* = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{x}'\mathbf{x}}\mathbf{x} = \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}.$$

8 ANOVA results

8.1 Sum of squares

In Chapter 2, we have studied the following partition of total sum of squares:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \;=\; \sum(\hat{Y}_i - \bar{Y})^2 \;+\; \sum(Y_i - \hat{Y}_i)^2.$$

We denote this partition as follows:

$$\text{SSTo} \;=\; \text{SSR} \;+\; \text{SSE}.$$

The sums of squares above can be expressed in matrix notation.

$$\begin{aligned}
\text{SSTo} &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 = \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} Y_i)^2 \\
&= \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{1}\mathbf{1}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y} \\
\text{SSE} &= \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\
&= \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{H}\mathbf{Y} \\
\text{SSR} &= \text{SSTo} - \text{SSE} \\
&= \mathbf{Y}'\mathbf{H}\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y},
\end{aligned}$$

where $\mathbf{J} = \mathbf{1}\mathbf{1}'$.

8.2 Sum of squares as quadratic forms

A quadratic form is defined as:

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}Y_iY_j, \qquad \text{where } a_{ij} = a_{ji}.$$

The $n \times n$ symmetric matrix $\mathbf{A}$ is called the *matrix of the quadratic* form.

Example 5.7. The quadratic equation $5Y_1^2 + 6Y_1Y_2 + 4Y_2^2$ can be rewritten by

$$\begin{bmatrix} Y_1 & Y_2 \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \mathbf{Y}'\mathbf{A}\mathbf{Y}.$$

$\|$

Using $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{Y}$ and $\hat{\boldsymbol{\beta}}'\mathbf{X}' = \mathbf{Y}'\mathbf{H}$, we have

$$\begin{aligned} \text{SSTo} &= \mathbf{Y}'\left[\mathbf{I} - \tfrac{1}{n}\mathbf{J}\right]\mathbf{Y} \\ \text{SSE} &= \mathbf{Y}'\left[\mathbf{I} - \mathbf{H}\right]\mathbf{Y} \\ \text{SSR} &= \mathbf{Y}'\left[\mathbf{H} - \tfrac{1}{n}\mathbf{J}\right]\mathbf{Y}. \end{aligned}$$

9 The geometry of ANOVA

Projecting the vector $\mathbf{Y}$ onto $\mathbf{1}$, we have the similar results to the case of the geometry of the least squares procedures.

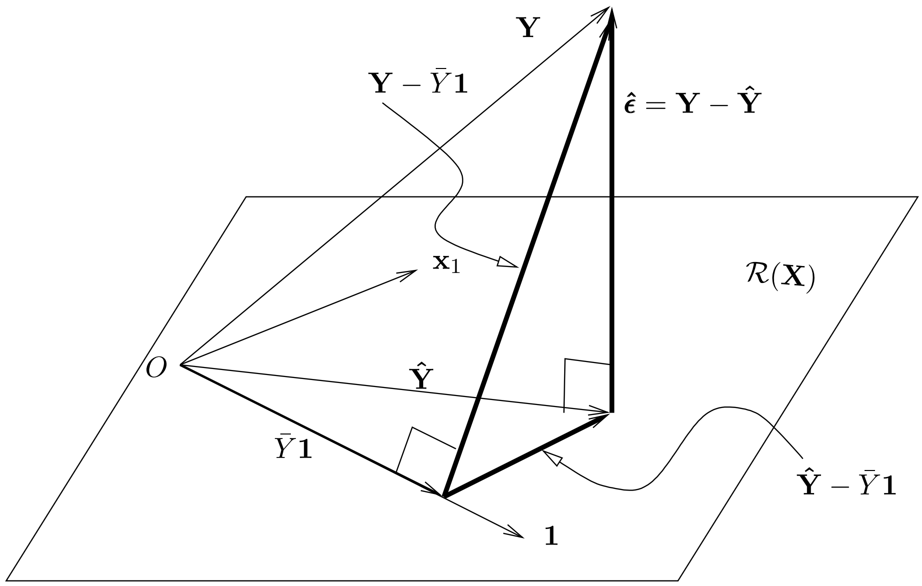
10 Inferences in regression analysis

1. *Regression Coefficient.*

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}}) &= \text{Cov}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \mathbf{Y}\right) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \text{Cov}(\mathbf{Y}) \cdot \left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right\}' \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \sigma^2\mathbf{I} \cdot \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}) = \text{MSE} \cdot (\mathbf{X}'\mathbf{X})^{-1}$$

The geometry of ANOVA



2. *Mean Response at* X_h.

Let us define

$$\mathbf{x}_h = \begin{pmatrix} 1 \\ X_h \end{pmatrix} \text{ and } \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}.$$

The fitted value at X_h in matrix notation is

$$\hat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = [1 \ X_h] \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \mathbf{x'}_h \hat{\boldsymbol{\beta}}.$$

Hence we have

$$\begin{aligned}
\mathrm{Cov}(\hat{Y}_h) = \mathrm{Var}(\hat{Y}_h) &= \mathrm{Cov}(\mathbf{x'}_h \hat{\boldsymbol{\beta}}) \\
&= \mathbf{x'}_h \,\mathrm{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{x}_h \\
&= \mathbf{x'}_h \sigma^2 (\mathbf{X'X})^{-1} \mathbf{x}_h \\
&= \sigma^2 \mathbf{x'}_h (\mathbf{X'X})^{-1} \mathbf{x}_h \\
\widehat{\mathrm{Cov}}(\hat{Y}_h) = \widehat{\mathrm{Var}}(\hat{Y}_h) &= \mathrm{MSE} \cdot \mathbf{x'}_h (\mathbf{X'X})^{-1} \mathbf{x}_h.
\end{aligned}$$

Thus, we have the distribution

$$\frac{\hat{Y}_h - E(\hat{Y}_h)}{\sqrt{\mathrm{Var}(\hat{Y}_h)}} \ \sim \ N(0,1) \ \text{ and } \ \frac{\hat{Y}_h - \mu_{Y_h}}{\sqrt{\widehat{\mathrm{Var}}(\hat{Y}_h)}} \ \sim \ t(n-p).$$

3. *Prediction of new observation,* $Y_{h(\text{new})}$

Recall that we have studied the following in §2.4:

$$\boxed{\mathrm{Var}(Y_{h(\text{new})}-\hat{Y}_h)=\mathrm{Var}(Y_{h(\text{new})})+\mathrm{Var}(\hat{Y}_h)=\sigma^2\Big\{1+\frac{1}{n}+\frac{(X_h-\bar{X})^2}{S_{xx}}\Big\}.}$$

Thus, the variance of $Y_{h(\text{new})}-\hat{Y}_h$ in matrix notation becomes

$$\mathrm{Var}(Y_{h(\text{new})}-\hat{Y}_h)=\sigma^2\big\{1+\mathbf{x}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h\big\}$$
$$\widehat{\mathrm{Var}}(Y_{h(\text{new})}-\hat{Y}_h)=\mathrm{MSE}\cdot\big\{1+\mathbf{x}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h\big\}.$$

It immediate from $E(Y_{h(\text{new})}-\hat{Y}_h)=0$ that

$$\frac{Y_{h(\text{new})}-\hat{Y}_h}{\sigma\sqrt{1+\mathbf{x}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h}}\ \sim\ N(0,1)$$
$$\frac{Y_{h(\text{new})}-\hat{Y}_h}{\sqrt{\mathrm{MSE}}\sqrt{1+\mathbf{x}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h}}\ \sim\ t(n-p).$$

Converting the above for $Y_{h(\text{new})}$, we have the prediction limits for μ_{Y_h}:

$$\boxed{\hat{Y}_h\pm t(1-\frac{\alpha}{2};n-p)\cdot\sqrt{\mathrm{MSE}\cdot\Big\{1+\mathbf{x}'_h(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_h\Big\}}.}$$

Example 5.8. Confidence limits of $\hat{Y}_h$. (See §2.3 of the handout).

It is immediate from

$$\mathbf{X}'\mathbf{X}=\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2\end{bmatrix},$$

that we have

$$(\mathbf{X}'\mathbf{X})^{-1}=\frac{1}{n\sum X_i^2-(\sum X_i)^2}\begin{bmatrix}\sum X_i^2 & -\sum X_i\\ -\sum X_i & n\end{bmatrix}=\frac{1}{nS_{xx}}\begin{bmatrix}\sum X_i^2 & -\sum X_i\\ -\sum X_i & n\end{bmatrix}$$

because $n\sum X_i^2-(\sum X_i)^2=n\big(\sum X_i^2-n\bar{X}^2\big)=nS_{xx}$. Thus, we have

$$\begin{aligned}\mathrm{Cov}(\hat{Y}_h)&=\sigma^2[1\ X_h]\frac{1}{nS_{xx}}\begin{bmatrix}\sum X_i^2 & -\sum X_i\\ -\sum X_i & n\end{bmatrix}\begin{bmatrix}1\\ X_h\end{bmatrix}\\ &=\frac{\sigma^2}{nS_{xx}}[1\ X_h]\begin{bmatrix}\sum X_i^2 & -n\bar{X}\\ -n\bar{X} & n\end{bmatrix}\begin{bmatrix}1\\ X_h\end{bmatrix}\\ &=\frac{\sigma^2}{nS_{xx}}\Big\{\sum X_i^2-2X_h(n\bar{X})+nX_h^2\Big\}.\end{aligned}$$

It follows from $\sum X_i^2 = S_{xx} + n\bar{X}^2$ that

$$\begin{aligned}
\mathrm{Cov}(\hat{Y}_h) = \mathrm{Var}(\hat{Y}_h) &= \frac{\sigma^2}{nS_{xx}}\Big\{S_{xx} + n\bar{X}^2 - 2X_h(n\bar{X}) + nX_h^2\Big\} \\
&= \frac{\sigma^2}{nS_{xx}}\Big\{S_{xx} + n(X_h - \bar{X})^2\Big\} \\
&= \sigma^2\Big\{\frac{1}{n} + \frac{(X_h - \bar{X})^2}{S_{xx}}\Big\}.
\end{aligned}$$

‖

References

Cochran, W. G. (1934). The distribution of quadratic forms in a normal system with applications to the analysis of variance. *Proceedings of the Cambridge Philosophical Society*, 30:178–191.

Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Wadsworth, Inc., 2nd edition.

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics*, volume 2. Charles Griffin, fourth edition.

Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill, New York, 5th edition.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley & Sons, New York.