SLR in the Matrix Form

If you are not familiar with matrix algebra,
please read KNNL Sections 5.1-5.7.

Matrix form

- Matrices and vectors will be written in **bold face** type.
- Subscripts will indicate the *dimension* of the matrix. For example,

$$\boldsymbol{A}_{\boldsymbol{3X2}} = \begin{bmatrix} 1 & 0.2 \\ 1 & 3.4 \\ 1 & 2.1 \end{bmatrix}$$

is a 3-row by 2-column matrix.

- Dimension subscripts will only be used when needed for clarity.

Warm up exercise

$$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} \qquad \underset{2\times 2}{\mathbf{B}} = \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix}$$

A) $\boldsymbol{AB} = \begin{bmatrix} 8 & 30 \\ 2 & 28 \end{bmatrix}$

B) $\boldsymbol{AB} = \begin{bmatrix} 38 \\ 28 \end{bmatrix}$

C) $\boldsymbol{AB} = \begin{bmatrix} 38 \\ 28 \end{bmatrix}$

D) $\boldsymbol{AB} = \begin{bmatrix} 33 & 52 \\ 21 & 32 \end{bmatrix}$

Warm up exercise

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{y_i} \end{bmatrix}$$

$$\boldsymbol{Y'Y} = \boldsymbol{\Sigma y_i^2}$$

$$\boldsymbol{Y'} \begin{bmatrix} \mathbf{1..1} \\ \mathbf{1..1} \\ \mathbf{1..1} \end{bmatrix} \boldsymbol{Y} = (\Sigma y_i)^2$$

SST $= \boldsymbol{\Sigma y_i^2} - (\Sigma y_i)^2 = \boldsymbol{Y'Y} - \boldsymbol{Y'} \begin{bmatrix} \mathbf{1..1} \\ \mathbf{1..1} \\ \mathbf{1..1} \end{bmatrix} \boldsymbol{Y}$

The SLR Model in Scalar Form

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{where} \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

Consider now writing an equation for each observation:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \quad &\vdots \quad \vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned}$$

The SLR Model in Matrix Form

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In matrix notation, the simple linear regression model is written as,

$$\boldsymbol{Y}_{n\times 1} = \boldsymbol{X}_{n\times 2}\boldsymbol{\beta}_{2\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

Or more simply as,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\boldsymbol{Y}$ is the *response vector*
- $\boldsymbol{X}$ is called the *design matrix*
- $\boldsymbol{\beta}$ is the *parameter vector*
- $\boldsymbol{\varepsilon}$ is the *error vector*

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Example (Flavor deterioration). The results shown below were obtain in a small-scale experiment to study the relation between F of storage temperature (X) and number of weeks before flavor deterioration of a food product begins occur (Y).

Observation	1	2	3	4	5
Temp (X)	8	4	0	-4	-8
Week (Y)	7.8	9	10.2	11	11.7

Write the matrix notation for X and Y. $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

```
week<-c(7.8,9,10.2,11,11.7)
week
```

```
[1]  7.8  9.0 10.2 11.0 11.7
```

```
y<-as.matrix(week)
colnames(y)<-c("week")
y
```

```
     week
[1,]  7.8
[2,]  9.0
[3,] 10.2
[4,] 11.0
[5,] 11.7
```

```
temp<-c(8,4,0,-4,-8)
Intercept<-rep(1,5)
x<-cbind(Intercept,temp)
x
```

```
     Intercept temp
[1,]         1    8
[2,]         1    4
[3,]         1    0
[4,]         1   -4
[5,]         1   -8
```

Example (Flavor deterioration). The results shown below were obtain in a small-scale experiment to study the relation between F of storage temperature (X) and number of weeks before flavor deterioration of a food product begins occur (Y).

Observation	1	2	3	4	5
Temp (X)	8	4	0	-4	-8
Week (Y)	7.8	9	10.2	11	11.7

Use the matrix notation to compute SST.

$$SST = \Sigma(Y-\bar{Y})^{\mathbf{2}} = \Sigma \mathrm{Y}^2 - \frac{(\Sigma \mathrm{Y})^2}{\mathrm{n}} = \boldsymbol{Y^tY} - \frac{\mathbf{1}}{\boldsymbol{n}}\boldsymbol{Y^tJY} = 9.752$$

Where $\boldsymbol{J}$ is the n by n matrix of 1s $\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$

$$Y = \begin{array}{c} \text{week} \\ \begin{pmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{pmatrix} \end{array}$$

```
J<-matrix(1,nrow=5,ncol=5)
SST<-t(y)%*%y-(1/5)*t(y)%*%J%*%y
```

Introducing the Variance-Covariance Matrix for a random multivariable, U

- $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U_1} \\ \boldsymbol{U_2} \\ \boldsymbol{U_3} \\ \dots \\ \boldsymbol{U_n} \end{bmatrix}$ is a multivariate. The mean, $\mathrm{E}\{\boldsymbol{U}\} = \begin{bmatrix} \boldsymbol{E(U_1)} \\ \boldsymbol{E(U_2)} \\ \boldsymbol{E(U_3)} \\ \dots \\ \boldsymbol{E(U_n)} \end{bmatrix}$

- The variance,

$$Var(U_i) = \sigma^2(U_i) = E\left[\left(U_i - E(U_i)\right)\left(U_i - E(U_i)\right)'\right] = E\left(U_i^2\right) - [E(U_i)]^2$$

- The covariance of two random multivariate, U and V, describes the linear relationship between $U\ and\ V$

$$\mathrm{Cov}\left(U_i, U_j\ \right) = \sigma\left(U_i, U_j\right) = E[\left(U_i - E(U_i)\right)\left(U_j - E\left(U_j\right)\right] = E\left(U_i U_j\right) - E(\mathrm{U}_i)\mathrm{E}(\mathrm{U_j})$$

- The variance-Covariance Matrix is a n-by-n matrix consisting both variance and covariance in the multivariate.

$$\Sigma\{\boldsymbol{U}\}_{nXn} = \begin{bmatrix} Var(U_i), \mathrm{Cov}\left(U_i, U_j\ \right) \\ \mathrm{Cov}\left(U_j, U_i\ \right), Var\left(U_j\right) \end{bmatrix}$$

Features on Normal Variable

- Usually, independent variables are always uncorrelated, but uncorrelated variables are not necessarily independent. That is, correlation only measures dependence in the linear dimension.

- Except when the variables follow Normal distribution, in which case correlation and dependence are the same.

- If variables are (completely) uncorrelated, their covariance is 0.

- The variance-covariance matrix of uncorrelated variables will therefore be a diagonal matrix , since all the covariances are 0.

$$\Sigma\{\boldsymbol{U}\}_{nXn} = \begin{bmatrix} Var(U_i), 0 \\ 0, Var\left(U_j\right) \end{bmatrix}$$

More General Notations

Let $U \sim N(E(U), \Sigma(U))$ be a multivariate normal vector,

and let $V = c + DU$ be a linear transformation of $\boldsymbol{U}$

where c is a vector of constants and D is a matrix of constants.

Then $V \sim N(c + D\mu, D\Sigma D^t)$.

The Variance-Covariance Matrix for the Random Error, $\Sigma\{\varepsilon\}$

$$\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \dots\\ \varepsilon_n\end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix}0\\0\\0\\ \dots\\0\end{bmatrix}, \begin{bmatrix}\sigma^2\ 0\ 0\ \dots 0\\ 0\ \sigma^2\ 0\ \dots 0\\ 0\ 0\ \sigma^2\ \dots 0\\ \dots\\ 0\ 0\ 0\ \dots \sigma^2\end{bmatrix}\right)$$

$$\Sigma\{\varepsilon\}_{nXn} = \begin{bmatrix}\sigma^2\ 0\ 0\ \dots 0\\ 0\ \sigma^2\ 0\ \dots 0\\ 0\ 0\ \sigma^2\ \dots 0\\ \dots\\ 0\ 0\ 0\ \dots \sigma^2\end{bmatrix} = \sigma^2 \mathrm{I}_{nXn}$$

This is true when the random errors are
- independent,
- have a mean of 0, and
- a constant variance

$$\begin{bmatrix}Y_1\\ Y_2\\ Y_3\\ \dots\\ Y_n\end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix}X_1\beta\\ X_2\beta\\ X_3\beta\\ \dots\\ X_n\beta\end{bmatrix}, \begin{bmatrix}\sigma^2\ 0\ 0\ \dots 0\\ 0\ \sigma^2\ 0\ \dots 0\\ 0\ 0\ \sigma^2\ \dots 0\\ \dots\\ 0\ 0\ 0\ \dots \sigma^2\end{bmatrix}\right)$$

$$\Sigma\{Y\}_{nXn} = \Sigma\{\varepsilon\}_{nXn}$$

- $Y = X\beta + \varepsilon = X\beta + I\varepsilon$, and
- No assumption violation

Least Squares Parameter Estimation

We want to minimize the sum of residuals.

To find the solution, set the derivative with respect to the vector $\boldsymbol{\beta}$ equal to a zero vector and

solve:

$$\frac{d}{d\boldsymbol{\beta}}(\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon}) = \frac{d}{d\boldsymbol{\beta}}((\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}))$$

$$= -2\boldsymbol{X}^t(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

(Partially) Solving for $\boldsymbol{\beta}$ yields the so-called *Normal Equations* :

$$-2\boldsymbol{X}^t(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}$$

$$\boldsymbol{X}^t\boldsymbol{Y} = \boldsymbol{X}^t\boldsymbol{X}\boldsymbol{\beta}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}(\boldsymbol{X}^t\boldsymbol{Y}) = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} (\bar{\boldsymbol{Y}}) - \frac{(\bar{X})SS_{XY}}{SS_X} \\ \frac{SS_{XY}}{SS_X} \end{bmatrix}$$

The details

$$\boldsymbol{\beta} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}(\boldsymbol{X}^t\boldsymbol{Y}) = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

$$X^tX = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}$$

$$(X^tX)^{-1} = \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & \mathrm{n} \end{bmatrix} = \frac{1}{nSS_X} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & \mathrm{n} \end{bmatrix}$$

$$X^tY = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}$$

Plug these into the equation for b:

$$\mathbf{b} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}(\boldsymbol{X}^t\boldsymbol{Y})$$

$$= \frac{1}{nSS_X}\begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & \mathrm{n} \end{bmatrix}\begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}$$

$$= \frac{1}{nSS_X}\begin{bmatrix} \left(\Sigma X_i^2\right)(\Sigma Y_i) - (\Sigma X_i)(\Sigma X_i Y_i) \\ -(\Sigma X_i)(\Sigma Y_i) + n(\Sigma X_i Y_i) \end{bmatrix}$$

$$= \frac{1}{SS_X}\begin{bmatrix} \left(\Sigma X_i^2\right)(\bar{\boldsymbol{Y}}) - (\bar{X})(\Sigma X_i Y_i) \\ -n(\bar{\boldsymbol{X}})(\bar{\boldsymbol{Y}}) + (\Sigma X_i Y_i) \end{bmatrix}$$

$$= \frac{1}{SS_X}\begin{bmatrix} \left(\Sigma X_i^2\right)(\bar{\boldsymbol{Y}}) - \bar{Y}\left(n\,\bar{X}^2\right) + \bar{X}(n\,\bar{X}\,\bar{Y}\,) - (\bar{X})(\Sigma X_i Y_i) \\ SS_{XY} \end{bmatrix}$$

$$= \frac{1}{SS_X}\begin{bmatrix} SS_X(\bar{\boldsymbol{Y}}) - (\bar{X})SP_{XY} \\ SS_{XY} \end{bmatrix} = \begin{bmatrix} (\bar{\boldsymbol{Y}}) - \frac{(\bar{X})SS_{XY}}{SS_X} \\ \frac{SS_{XY}}{SS_X} \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$ Finally!

Example (Flavor deterioration). The results shown below were obtain in a small-scale experiment to study the relation between F of storage temperature (X) and number of weeks before flavor deterioration of a food product begins occur (Y).

i	1	2	3	4	5
Xi	8	4	0	-4	-8
Yi	7.8	9	10.2	11	11.7

Compute estimators in matrix form

$$Y_{5X1} = \begin{array}{c} \text{week} \\ \begin{pmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{pmatrix} \end{array}$$

$$X_{5X2} = \begin{array}{cc} \text{Intercept} & \text{temp} \end{array} \begin{pmatrix} 1 & 8 \\ 1 & 4 \\ 1 & 0 \\ 1 & -4 \\ 1 & -8 \end{pmatrix}$$

```
xtx<-t(x)%*%x
xtxinv<-solve(xtx)
xtxinv
```

```
          Intercept    temp
Intercept       0.2 0.00000
temp            0.0 0.00625
```

```
xty<-t(x)%*%y
xty
```

```
           week
Intercept  49.7
temp      -39.2
```

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}(\boldsymbol{X'Y}) = \begin{bmatrix} 9.94 \\ -0.245 \end{bmatrix}$$

The Hat Matrix

$$\begin{aligned}\hat{Y} &= X\,b \\ &= X\,(X^tX)^{-1}X^tY \\ &= H\,Y\end{aligned}$$

where $H = X\,(X^tX)^{-1}X^t$ is called the *hat matrix* because it turns Y's into $\hat{Y}$'s.

The hat matrix will give us many useful diagnostic tools in MLR.

In the flavor example

$$\underset{5X5}{H} = \begin{pmatrix} 0.6 & 0.4 & 0.2 & 0.0 & -0.2 \\ 0.4 & 0.3 & 0.2 & 0.1 & 0.0 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ 0.0 & 0.1 & 0.2 & 0.3 & 0.4 \\ -0.2 & 0.0 & 0.2 & 0.4 & 0.6 \end{pmatrix}$$

The matrix H is symmetric and has the special property called **idempotent**:

$$HH = H$$

Predicted value, residual, SSE, and MSE

$$\hat{\boldsymbol{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \boldsymbol{X}\boldsymbol{b}$$

$$\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$$

$$= \begin{bmatrix} -0.18 \\ 0.04 \\ 0.26 \\ 0.08 \\ -0.20 \end{bmatrix}$$

$$\boldsymbol{SSE} = \boldsymbol{e}'\boldsymbol{e} = \begin{bmatrix} -0.18 & 0.04 & 0.26 & 0.08 & -0.2 \end{bmatrix} \begin{bmatrix} -0.18 \\ 0.04 \\ 0.26 \\ 0.08 \\ -0.20 \end{bmatrix}$$

$$= 0.148$$

$$\boldsymbol{MSE} = SSE/(n-p) = 0.\mathbf{0493}$$

The Matrix Form of the coefficients

$$\boldsymbol{b} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y}$$

And $\quad \boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$

Since b is a linear combination of Y, b follows Normal distribution.

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{b}) &= (X^tX)^{-1}X^t\,\boldsymbol{E}(\boldsymbol{Y}) \\ &= (X^tX)^{-1}X^t\,X\,\boldsymbol{\beta} = \boldsymbol{\beta} \end{aligned}$$

Therefore, $\boldsymbol{b}$ is an unbiased estimator of $\boldsymbol{\beta}$.

The variance-covariance matrix for the coefficient

$$\begin{aligned}\Sigma_{\{b\}} &= \left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]\sigma^2\boldsymbol{I}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]' \\ &= \sigma^2\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]\boldsymbol{I}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]' \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]' \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}$$

- In the last step, $\boldsymbol{X}^t\boldsymbol{X}$ and its inverse are both symmetric, therefore $((\boldsymbol{X}^t\boldsymbol{X})^{-1})^t = (\boldsymbol{X}^t\boldsymbol{X})^{-1}$
- $\sigma^2(\boldsymbol{X}^t\boldsymbol{X})^{-1}$ generally is **NOT** a diagonal matrix, because the estimates b_0 and b_1 are generally **not** independent of each other.
- $\Sigma\{b\}$ has a dimension of p by p, where p is the number of parameters.

Example (Flavor deterioration).

Compute the covariance matrix for $\boldsymbol{b}$

$$Y = \begin{array}{r} \text{week} \\ \begin{bmatrix} 7.8 \\ 9.0 \\ 10.2 \\ 11.0 \\ 11.7 \end{bmatrix} \end{array}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}(\boldsymbol{X}^t\boldsymbol{Y}) = \begin{array}{lr} & \text{week} \\ \text{Intercept} & 9.940 \\ \text{temp} & -0.245 \end{array}$$

$$\boldsymbol{\Sigma}\{\boldsymbol{b}\} = \boldsymbol{\sigma}^2 \ (\boldsymbol{X}^t\boldsymbol{X})^{-1} = \boldsymbol{MSE}(\boldsymbol{X}^t\boldsymbol{X})^{-1}$$

$$= 0.0493 \begin{pmatrix} 0.2 & 0.00000 \\ 0.0 & 0.00625 \end{pmatrix} = \begin{pmatrix} \boxed{0.009866667} & 0.000000000 \\ 0.000000000 & \boxed{0.0003083333} \end{pmatrix}$$

$$X = \begin{array}{rr} \text{Intercept} & \text{temp} \\ 1 & 8 \\ 1 & 4 \\ 1 & 0 \\ 1 & -4 \\ 1 & -8 \end{array}$$

- The diagonal elements in $\boldsymbol{\Sigma}\{\boldsymbol{b}\}$ are $s^2\{b_0\}$ and $s^2\{b_1\}$
- The non-diagonal elements are generally not 0 due to dependency.

The Matrix Form of the Mean Response at One X

To estimate the mean response at $\boldsymbol{X_h}$, define a matrix form $\underset{2\times1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_h \end{bmatrix}$

$$\hat{Y}_h = \mathbf{X}_h'\mathbf{b} = [1 \quad X_h]\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1 X_h]$$

$$\boldsymbol{\Sigma}\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\} = \boldsymbol{X}_{\boldsymbol{h}}'\boldsymbol{\Sigma}\{\mathbf{b}\}\mathbf{X_h}$$

Example (Flavor deterioration): find the point estimator and standard error of $\hat{Y}_h$ when $\mathrm{X_h} = -6$

$$\hat{Y}_h = \mathbf{X}_h'\mathbf{b} = [1, -6]\begin{bmatrix} 9.94 \\ -0.245 \end{bmatrix} = 11.41$$

$$\boldsymbol{\Sigma}\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\} = \boldsymbol{X}_{\boldsymbol{h}}'\boldsymbol{\Sigma}\{\mathbf{b}\}\mathbf{X_h} = [1 \quad -6]\begin{bmatrix} 0.00986 & 0 \\ 0 & 0.00031 \end{bmatrix}\begin{bmatrix} 1 \\ -6 \end{bmatrix}$$

$$= 0.021$$

$$= \boldsymbol{s}^2\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\}$$

- The variance covariance matrix of the mean response has a dimension of 1 by 1, and reduces to $\boldsymbol{s}^2\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\}$
- Note that the X_h matrix differs from the design matrix!

The Matrix Form of the Mean Response at Multiple Xs

To estimate the mean response at m multiple points, define a 2 by m matrix form for X_h

Example (Flavor deterioration): find the point estimator and standard error of $\hat{Y}_h$ when $X = -6 \text{ and } -5$

$$X_h = \begin{bmatrix} 1 & 1 \\ -6 & -5 \end{bmatrix} \qquad \hat{Y}_h = \mathbf{X}_h'\mathbf{b} = \begin{bmatrix} 1 & -6 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 9.94 \\ -0.245 \end{bmatrix} = \begin{bmatrix} 11.41 \\ 11.17 \end{bmatrix}$$

$$\mathbf{\Sigma}\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\} = \boldsymbol{X}_{\boldsymbol{h}}'\mathbf{\Sigma}\{\mathbf{b}\}\mathbf{X_h} = \begin{bmatrix} 1 & -6 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 0.00986 & 0 \\ 0 & 0.00031 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -6 & -5 \end{bmatrix} = \begin{bmatrix} 0.021 & * \\ * & 0.018 \end{bmatrix}$$

- The diagonal values, $\boldsymbol{s}^2\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\}$ = 0.021 and 0.018 for the X=-6 and -5, respectively.
  - Use the $\boldsymbol{s}^2\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\}$ to compute the confidence interval for the mean response: $\hat{\boldsymbol{Y}}_{\boldsymbol{h}} \pm \boldsymbol{ts}\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\}$
  - Use the $\boldsymbol{s}^2\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\}$ to compute the variance (and the CI) for the single response: $\boldsymbol{s}^2\{\boldsymbol{pred}\} = \boldsymbol{s}^2 + \boldsymbol{s}^2\{\hat{\boldsymbol{Y}}_{\boldsymbol{h}}\}$
- The off-diagonal values, or the “*” part is not applicable under the assumption of independence.
- $\Sigma\{\hat{Y}_h\}$ has a dimension of m by m, where m is the number of X levels to predict.