Review of Multiple Linear Regression

Data $(X_{i1}, X_{i2}, \ldots, X_{i,p-1}, Y_i)$, $i = 1, 2, \ldots, n$

Model Equation and Assumptions

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{p-1} X_{i,p-1} + \epsilon_i$$

- $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$
- $\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$ and σ^2 are unknown parameters
- X_{ij}'s are known constants

Two cases:

1. $p-1$ different predictors

2. some of the predictors are functions of the others

   (a) polynomial regression

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

   (b) interaction effects

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$$

General Linear Model in Matrix Terms

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \mathbf{X}_{n\times p} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{bmatrix}$$

$$\boldsymbol{\beta}_{p\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \qquad \boldsymbol{\epsilon}_{n\times 1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Model:

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

Assumptions:

- $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$
- $\boldsymbol{\beta}$ and σ^2 are unknown parameters
- $\mathbf{X}$ is a $(n \times p)$ matrix of fixed known constants

Random Vectors and Matrices

- A random vector is a vector of random variables, e.g. $\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix}'$
- The **expected value** of $\mathbf{Y}$ is a vector, denoted by $E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) & E(Y_2) & \dots & E(Y_n) \end{bmatrix}'$.
- **The usual rules for expectation still work:** Suppose $\mathbf{V}$ and $\mathbf{W}$ are random vectors and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrices of constants. Then

$$E(\mathbf{AV} + \mathbf{BW} + \mathbf{C}) = \mathbf{A}E(\mathbf{V}) + \mathbf{B}E(\mathbf{W}) + \mathbf{C}$$

Regression Example: Find $E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})$

$$E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = E(\mathbf{X}\boldsymbol{\beta}) + E(\boldsymbol{\epsilon}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{0} = \mathbf{X}\boldsymbol{\beta}$$

Variance-Covariance Matrix of a Random Vector

For a random vector $\mathbf{Z}_{n\times 1}$ define $Var(\mathbf{Z}) =$

$$\begin{bmatrix} Var(Z_1) & Cov(Z_1,Z_2) & \dots & Cov(Z_1,Z_n) \\ Cov(Z_2,Z_1) & Var(Z_2) & \dots & Cov(Z_2,Z_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Z_n,Z_1) & Cov(Z_n,Z_2) & \dots & Var(Z_n) \end{bmatrix}$$

where $Cov(Z_i,Z_j) = E\Big[(Z_i - E(Z_i))(Z_j - E(Z_j))\Big] = Cov(Z_j,Z_i)$. It is a symmetric $(n \times n)$ matrix.

If Z_i and Z_j are independent, then $Cov(Z_i,Z_j) = 0$.

Correlation Coefficient: The correlation between to random variables is defined as

$$Cor(Z_i,Z_j) = \frac{Cov(Z_i,Z_j)}{\sqrt{Var(Z_i)Var(Z_j)}}$$

Rules Variance-Covariance Matrices

- Remember: if V is a r.v. and a, b are constant terms, then

$$Var(aV+b) = Var(aV) = a^2Var(V)$$

- Suppose now that $\mathbf{V}$ is a random vector and $\mathbf{A}$, $\mathbf{B}$ are matrices of constants. Then

$$Var(\mathbf{AV}+\mathbf{B}) = Var(\mathbf{AV}) = \mathbf{A}Var(\mathbf{V})\mathbf{A}'$$

- More generally

$$Cov(\mathbf{AV}+\mathbf{C},\mathbf{BW}+\mathbf{D}) = \mathbf{A}Cov(\mathbf{V},\mathbf{W})\mathbf{B}'$$

Regression Analysis: Find $Var(\mathbf{Y}) = Var(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon})$

$$Var(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}) = Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_{n\times n}$$

Off-diagonal elements are zero because the ϵ_i's, and hence the Y_i's, are independent.

Least Squares Estimates:

$$\mathbf{b}_{p\times 1} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Fitted Values:

$$\begin{aligned}
\hat{\mathbf{Y}}_{n\times 1} &= \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_{11} + \ldots + b_{p-1} X_{1,p-1} \\ b_0 + b_1 X_{21} + \ldots + b_{p-1} X_{2,p-1} \\ \vdots \\ b_0 + b_1 X_{n1} + \ldots + b_{p-1} X_{n,p-1} \end{bmatrix} \\
&= \mathbf{X}\mathbf{b} \\
&= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{H}\mathbf{Y}
\end{aligned}$$

Residuals:

$$\begin{aligned}
\mathbf{e}_{n\times 1} &= \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\mathbf{b} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= (\mathbf{I} - \mathbf{H})\mathbf{Y}
\end{aligned}$$

ANalysis Of VAriance

Formulas are exactly the same. Remember

$$\begin{aligned}
SSTO &= SSR + SSE \\
\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2
\end{aligned}$$

but their degrees of freedom (df) change:

- $SSTO$ still has $n-1$ df
- SSR now has $p-1$ because of the p param's in $\hat{Y}_i$
- SSE therefore has $n-p$ df

ANOVA Table for MLR:

Source variat.	Sum of Squares (SS)	df	mean SS
Regr.	$SSR = \sum_i (\hat{Y}_i - \bar{Y})^2$	$p-1$	$\frac{SSR}{p-1}$
Error	$SSE = \sum_i (Y_i - \hat{Y}_i)^2$	$n-p$	$\frac{SSE}{n-p}$
Total	$SSTO = \sum_i (Y_i - \bar{Y})^2$	$n-1$	

Overall F-Test for Regression Relation

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$

H_A : not all β_j $(j = 1, \ldots, p-1)$ equal zero.

H_0 states that all predictors $X_1, \ldots, X_{p-1}$ are useless (no relation between Y and the set of X variables), whereas H_A says that at least one is useful.

Test Statistic

$$F^* = \frac{MSR}{MSE}$$

Rejection Rule: reject H_0, if $F^* > F(1-\alpha; p-1, n-p)$

Note: when $p-1 = 1$, this is the F-test for $H_0 : \beta_1 = 0$ in the SLR.

Coefficient of Multiple Determination: it's the same as in SLR's,

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

It measures the relative reduction in the total variation $(SSTO)$ due to the MLR.

Inferences about Regression Parameters

Since with $\mathbf{C}_{p\times n} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ we can write

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} c_{11} & \ldots & c_{1n} \\ \vdots & \vdots & \vdots \\ c_{p1} & \ldots & c_{pn} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

Thus, every element of $\mathbf{b}$ is a linear combination of the Y's and is therefore a normal r.v.

Again

$$E(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) = \boldsymbol{\beta}$$

Thus $\mathbf{b}$ is an unbiased estimator for $\boldsymbol{\beta}$. Moreover

$$Var(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

This means that for any $k = 0, 1, \ldots, p-1$ we have

$$b_k \sim N\left(\beta_k, \sigma^2 \cdot \left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{k+1,k+1}\right)$$

where $[\cdot]_{jj}$ is the jth diagonal element of the matrix.

Thus

$$\frac{b_k - \beta_k}{\sqrt{\sigma^2 \cdot \left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{k+1,k+1}}} \sim N(0,1)$$

and because the MSE now has $df = n - p$

$$\frac{b_k - \beta_k}{\sqrt{MSE \cdot \left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{k+1,k+1}}} \sim t(n-p)$$

Using this we can construct tests and CI's for each individual β_k

Test Statistic:

$$t^* = \frac{b_k}{\sqrt{MSE \cdot \left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{k+1,k+1}}}$$

Rejection Rule: reject H_0 if $t^* > t(1-\alpha/2; n-p)$

- $(1-\alpha)100\%$ **CI for the parameter** β_k

$$b_k \pm t(1-\alpha/2; n-p)\sqrt{MSE \cdot \left[(\mathbf{X}'\mathbf{X})^{-1}\right]_{k+1,k+1}}$$

- $(1-\alpha)100\%$ **CI for the mean** of Y at $\mathbf{X}_h = (1\ X_{h1}\ X_{h2}\ \dots\ X_{h,p-1})'$

The point estimate of $E(Y_h) = \mathbf{X}_h'\boldsymbol{\beta}$ is

$$\widehat{E}(Y_h) = \hat{Y}_h = \mathbf{X}_h'\mathbf{b}$$

Since $\mathbf{b} \sim N\left[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right]$, it follows that

$$E(\widehat{E}(Y_h)) = \mathbf{X}_h' E(\mathbf{b}) = \mathbf{X}_h'\boldsymbol{\beta}$$

(unbiased) and

$$Var(\widehat{E}(Y_h)) = \mathbf{X}_h'\ Var(\mathbf{b})\ \mathbf{X}_h = \sigma^2 \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h$$

The CI for $\mathbf{X}_h'\boldsymbol{\beta}$ is constructed in the usual manner:

$$\mathbf{X}_h'\mathbf{b} \pm t(1-\alpha/2; n-p)\sqrt{MSE \cdot \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h}$$