Simple Linear Regression in Matrix Form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2), \quad i = 1, \cdots, n$$

Response vector $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$

Design Matrix (predictor matrix) $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{pmatrix}$

Coefficient Vector $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

Error Vector $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

$$\underset{n\times 2}{X} \cdot \underset{2\times 1}{\beta} = (\beta_0 + \beta_1 x_i)_{i=1,\cdots,n}$$ matrix product

Linear Model: $Y = X\beta + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I_n)$.

$$Var(\varepsilon) = (Cov(\varepsilon_i, \varepsilon_j)) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix} = \sigma^2 I_n, \quad Cov(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$$

Least Square Criterion

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \qquad \frac{\partial Q}{\partial \beta_0} = 0, \ \frac{\partial Q}{\partial \beta_1} = 0$$

$$Q(\beta) = (Y - X\beta)'(Y - X\beta) \qquad \frac{\partial Q}{\partial \beta} = 0$$

Least Square Estimator

$$\hat{\beta}_{LS} = (X'X)^{-1} X' Y = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$
$$= AY$$

Fitted value $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad i=1,\cdots,n$

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

Hat matrix H : symmetric and idempotent $\Rightarrow$ H projection matrix

$$H = H', \quad H^2 = H \qquad \Rightarrow \text{rank}(H) = \text{trace}(H)$$

Residual: $\hat{e} = Y - \hat{Y} = (I_n - H)Y$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \hat{e}'\cdot\hat{e} = Y'(I_n - H)\cdot Y$$

$$DF(SSE) = \text{rank}(I_n - H) = \text{trace}(I_n - H)$$
$$= \text{trace}(I_n) - \text{trace}(H) = n-2$$

---

1. # Show that $\text{trace}(H) = 2$, $H\cdot J_n = J_n$,

$\hat{Y} = HY$, $\quad \hat{Y} = X\cdot\hat{\beta} = \begin{bmatrix}1 & x_1\\ \vdots & \vdots \\ 1 & x_n\end{bmatrix}\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{bmatrix}$ $\qquad \hat{\beta}_0 = \sum_{i=1}^{n} k_i y_i \quad k_i = \frac{1}{n} - \bar{x}c_i$

$\hat{\beta}_1 = \sum_{i=1}^{n} c_i y_i, \quad c_i = \frac{x_i - \bar{x}}{S_{xx}}$

<1>

$$= \begin{bmatrix}1 & x_1\\ \vdots & \vdots \\ 1 & x_n\end{bmatrix}\begin{bmatrix}k_1 & \cdots & k_n\\ c_1 & \cdots & c_n\end{bmatrix}\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

$$= \begin{bmatrix} x_i\cdot c_j + k_j\end{bmatrix}\cdot\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix} = H\cdot Y$$

i.e. $H = (h_{ij})_{n\times n}$, $\quad h_{ij} = x_i c_j + k_j$,

① $h_{ij} = h_{ji}$ $\qquad x_i c_j + k_j = x_i\cdot\frac{x_j - \bar{x}}{S_{xx}} + \left(\frac{1}{n} - \bar{x}c_j\right)$

H is symmetric $\qquad = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} = x_j c_i + k_i$

② $\text{trace}(H) = \sum_{i=1}^{n} h_{ii} = \sum_{i=1}^{n}(x_i c_i + k_i) = \sum_{i=1}^{n}\left(x_i c_i + \frac{1}{n} - \bar{x}c_i\right)$

$$= \sum_{i=1}^{n}\left(\frac{1}{n}\right) + \sum_{i=1}^{n}\frac{c_i(x_i - \bar{x})}{S_{xx}} = 1 + 1 = 2.$$

③ $\sum_{j=1}^{n} h_{ij} = \sum_{j=1}^{n}\left[(x_i - \bar{x})c_j + \frac{1}{n}\right] = \sum_{i=1}^{n}\left(\frac{1}{n}\right) = 1 \quad \Rightarrow H\cdot J_n = J_n$

<2>

or $\text{trace}(H)$

$= \text{trace}(X(X'X)^{-1}X')$

$= \text{trace}((X'X)^{-1}X'X) = \text{trace}(I_2) = 2$

2. Show that $Cov(\hat{\beta}, \hat{e}) = 0$

$\hat{\beta} = AY = (X'X)^{-1}X'Y$

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

$$Cov(\hat{\beta}, \hat{e}) = Cov(AY, (I-H)Y)$$
$$= A \cdot Var(Y) \cdot (I-H)'$$
$$= \sigma^2 \cdot A \cdot (I-H) = \sigma^2 (A - AH)$$

$$AH = \underline{(X'X)^{-1}X'} \cdot X \cdot (X'X)^{-1} \cdot X' = (X'X)^{-1}X' = A$$

$\therefore Cov(\hat{\beta}, \hat{e}) = 0$

$\hat{\beta}, \hat{e}$ are both normally distributed $\Big\}$ $\hat{\beta}$ is independant of $\hat{e}$

i.e. $\hat{\beta} \perp\!\!\!\perp \hat{e}$

$SSR = \hat{\beta}_1^2 \cdot S_{xx}$ $\quad$ $SSE = \hat{e}' \cdot \hat{e}$ $\quad$ $\Rightarrow SSR \perp\!\!\!\perp SSE$

$= Y'(H - \frac{1}{n}J_n)Y$ $\quad$ $= Y'(I-H)Y$

$$\bar{Y} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} y_i \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \vdots \\ \bar{y} \end{pmatrix} = \frac{1}{n} J_n \cdot Y, \quad J_n = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = (\hat{Y} - \bar{Y})'(\hat{Y} - \bar{Y})$$

$$= Y'(H - \frac{1}{n} J_n) Y$$

It can be shown that $H - \frac{1}{n} J_n$ is symmetric and idempotent.

$$\Rightarrow DF(SSR) = Rank(H - \frac{1}{n} J_n) = trace(H - \frac{1}{n} J_n) = 2 - 1 = 1.$$

$$SSTO = \sum_{i=1}^{n} (y_i - \bar{y})^2 = (Y - \bar{Y})'(Y - \bar{Y}) = Y'(I_n - \frac{1}{n} J_n) Y$$

$$\Longrightarrow DF(SSTO) = trace(I_n - \frac{1}{n} J_n) = n - 1.$$

In addition, we can show that $SSR \perp\!\!\!\perp SSE$

as $(I_n - H) \cdot (H - \frac{1}{n} J_n) = 0$.

$$\hat{\beta} = A \cdot Y, \quad E(\hat{\beta}) = A \cdot E(Y) = (X'X)^{-1} X' \cdot (X\beta) = \beta$$

$$Var(\hat{\beta}) = A \cdot Var(Y) \cdot A' = \sigma^2 \cdot (X'X)^{-1}$$

$$Y \sim N(X\beta, \sigma^2 I_n) \quad \Rightarrow \hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$

$$Var(\hat{\beta}) = \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_1, \hat{\beta}_0) & Var(\hat{\beta}_1) \end{pmatrix} = \sigma^2 \cdot \begin{pmatrix} \frac{\sum_i x_i^2}{n S_{xx}} & -\frac{\bar{x}}{S_{xx}} \\ \frac{-\bar{x}}{S_{xx}} & \frac{1}{S_{xx}} \end{pmatrix}$$

$$= \frac{\sigma^2}{S_{xx}} \begin{pmatrix} \overline{x^2} & -\bar{x} \\ \bar{x} & 1 \end{pmatrix}$$

§4. Multiple Linear Regression

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2), \quad i=1,\cdots,n$$

β_i : partial regression coefficient (other covariates held constant)

x_i : independent predictor variables

Additive effects — main effects

Interaction effects model

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} \cdot x_{2i} + \varepsilon_i \quad \text{(higher order model)}$$

Additive checking $H_0: \beta_{1,2} = 0$ vs $H_1: \beta_{12} \neq 0$.

Use Matrix Form $\quad Y = X\beta + \varepsilon$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & & x_{np} \end{pmatrix}_{n\times(p+1)} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{(p+1)\times 1}$$

Least Square Criterion $\quad Q(\beta) = (Y - X\beta)'(Y - X\beta)$

$$\frac{\partial Q}{\partial \beta} = 0 \quad \Rightarrow \quad \hat{\beta} = (X'X)^{-1}X'Y \sim N(\beta, \sigma^2 (X'X)^{-1})$$

Fitted Value $\hat{Y} = X\hat{\beta} = HY$

Residual $\hat{e} = Y - \hat{Y} = (I - H)Y$

$$SSTO = SSE + SSR$$

$$SSE = \sum_{i=1}^{n} e_i^2 = \hat{e}' \cdot \hat{e} = Y'(I-H)Y$$

$$SSR = Y'(H - \tfrac{1}{n}J_n)Y$$

$$DF(SSTo) = n-1$$

$$DF(SSR) = trace\left(H - \frac{1}{n} J_n\right) = (p+1) - 1 = p$$

$$DF(SSE) = trace(n - H) = n - (p+1)$$

where $trace(H) = trace(X(X'X)^{-1}X')$ $\quad X_{n\times(p+1)}$

$$= trace((X'X)^{-1}X'X) = p+1$$

$\therefore MSR = \dfrac{SSR}{p}$, $\quad MSE = \dfrac{SSE}{n-(p+1)}$, $\quad E[MSE] = \sigma^2$.

It can be shown that $\dfrac{SSR}{\sigma^2} \overset{\beta_1=\cdots=\beta_p=0}{\sim} \chi^2(p)$, $\dfrac{SSE}{\sigma^2} \sim \chi^2(n-(p+1))$

and $SSR \perp\!\!\!\perp SSE$.

$$E[SSE] = E(\hat{e}' \cdot \hat{e}) = E(Y'(I-H) \cdot Y)$$

$$= E[trace(Y'(I-H)Y)]$$

$$= E[trace((I-H)YY')]$$

$$= trace((I-H) \cdot (Var Y + (EY) \cdot (EY)'))$$

$$= trace((I-H) \cdot \sigma^2 I_n)$$

$$= \sigma^2 \cdot trace(I-H)$$

$$= \sigma^2 (n-p-1)$$

$(I-H) \cdot (EY)$
$= (I-H) \cdot X\beta = 0$

$$\therefore E[MSE] = E\left[\frac{SSE}{n-p-1}\right] = \sigma^2$$

Simple linear regression : $p=1$

$$E[MSR] = \sigma^2 + \beta_1^2 \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Multiple linear regression $(p \geq 2)$

$$E[MSR] = \sigma^2 + \frac{1}{p} \sum_{j=1}^{p} \left[\beta_j^2 \cdot \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \right]$$

(a) Under $H_0 : \beta_1 = \cdots = \beta_p = 0$ [vs H_1 : at least one $\beta_j \neq 0$]

$E[MSR] = \sigma^2$ (multiple test)

$$F = \frac{MSR}{MSE} \overset{H_0}{\sim} F(p, n-p-1)$$

Critical region $\mathcal{C} = \{ F > F_\alpha(p, n-p-1) \}$

(b) $R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$, $R^2_{adj} = 1 - \frac{SSE/(n-p-1)}{SSTO/(n-1)}$

$R^2 \uparrow$, $F \uparrow$

(c) Individual test (or partial coefficient test)

$H_0 : \beta_j = 0$ vs $H_1 : \beta_j \neq 0$

$se(\hat{\beta}_j) = \sqrt{D_{j+1} \cdot \hat{\sigma}^2} = \sqrt{D_{j+1} \cdot MSE}$

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \overset{H_0 : \beta_j = 0}{\sim} t(n-p-1)$$

$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$

let $A = (X'X)^{-1}$

it $(j+1)$-th diagonal $A_{j+1, j+1} = D_{j+1}$

ANOVA Table for Multiple Regression : $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$
$i = 1, \ldots, n$

Source of Variation	DF	Sum Squares	Mean Square	F
Regression	p	$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	$MSR = \frac{SSR}{p}$	$F = \frac{MSR}{MSE}$
Error	$n-(p+1)$	$SSE = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$	$MSE = \frac{SSE}{n-p-1}$	
Total	$n-1$	$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$		

<1>. For hypotheses: $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$ vs H_1: at least one $\beta_j \neq 0$

$$F = \frac{MSR}{MSE} \overset{H_0}{\sim} F(p, n-p-1)$$

Critical region $\mathcal{C} = \{ F > F_\alpha(p, n-p-1) \}$

<2>. partial test $H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$

$$t = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \overset{H_0: \beta_j = 0}{\sim} t(n-p-1), \qquad \hat{\beta}_j \pm t_{\frac{\alpha}{2}}(n-p-1) \cdot se(\hat{\beta}_j)$$

$\hat{Y} = X \cdot \hat{\beta}$, $\hat{\beta} = (X'X)^{-1} X'Y$, given that $X'X$ is of full rank or $|X'X| \neq 0$.

Otherwise there are redundant variables in the model.

Extra Topics in Multiple Regression

1. ① Type I Sum Square : Sequential Sum Square

x_1 x_2 x_3

$SS(x_1)$ $SS(x_2 | x_1)$ $SS(x_3 | x_1, x_2)$

② Type II Sum Square : Conditional Sum Square

x_1 x_2 x_3

$SS(x_1 | x_2, x_3)$ $SS(x_2 | x_1, x_3)$ $SS(x_3 | x_1, x_2)$

2. Multicollinearity among predictors

① Correlation matrix: $\rho_{ij} = Corr(x_i, x_j)$, $i, j = 1 \cdots p$ $\Big|$ $\{x_1, \cdots x_{j-1}, x_{j+1}, \cdots x_p\}$

② Partial R_j^2 : R_j^2 is the R^2 for regression x_j on $x_{(-j)}$

③ Variance of Inflation Factor (VIF)

$$VIF_j = \frac{1}{1-R_j^2},$$

If $VIF_j > 10$, then we think there exists partial collinearity in the data

$\sqrt{VIF_j} = 8$ means the standard error with the multicollinearity is 8 times bigger than the standard error in the independent case.

④ Multicollinearity will produce imprecise estimate for β, and partial t-test will fail to identify significant predictors as the standard error is inflated.

⑤ Condition Index : $ratio_i = \dfrac{\lambda_{max}}{\lambda_i}$, $i = 1, \cdots, p$, $\begin{pmatrix} >100 & \text{strong} \\ >10 & \text{weak} \end{pmatrix}$

where $\lambda_1, \cdots, \lambda_p$ are eigenvalues of a scaled matrix of $(X'X)^{-1}$ such that its first eigenvalue $= 1$.

3. Model Selection

① Forward : reduce SSE the most, add the most significant ones

② Backward : leave out the least significant predictors

③ stepwise

④ Mallow's C_p (address the issue of over-fitting)

$$C_p = \frac{SSE_p}{MSE_{full}} - N + 2p$$

⑤ AIC :

$$AIC(C_p) = -2 * \log L + 2(p+1)$$

4. Outlier or Influential Point Detection

① Cook's Distance : $D_i = \frac{e_i^2}{p \cdot SSE} \cdot \frac{h_{ii}}{(1-h_{ii})^2}$, e_i: residual of i-th observation

where h_{ii} is the i-th diagonal entry in the hat matrix $H = X(X'X)^{-1}X'$.

is also called the leverage.

For observation with $D_i > 1$, it will be identified as an influential point.

② Outlier Test:

H_0: not an outlier

$$T_i = \frac{e_i}{se_{(-i)} \cdot \sqrt{1-h_{ii}}} \sim t(n-p-2)$$

where $se_{(-i)}$ is the standard error of the data with i-th observation excluded.

For large n, could use $e_i \pm 2 \cdot se_{(-i)} \sqrt{1-h_{ii}}$ as an alternative for the test.