Part II: Multiple linear regression

L08 Multiple linear regression models

1. Multiple linear regression models with intercepts

   (1) Model

   $$\text{In Model } \; y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon, \; \epsilon \sim N(0, \sigma^2),$$

   the regression function $E(y) = \beta_0 + \cdots + \beta_k x_k$ contains multiple ($k > 1$) predictors, and $E(y) = (1, x_1, ..., x_k)\beta$ is a linear function of unknown parameter vector $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$.

   Hence the model is called a multiple linear regression model with intercept β_0.

   (2) Parameters
   $\beta \in R^p$ with $p = k + 1$, $\sigma^2 > 0$ and
   $E(y(x_0)) = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} = x_0'\beta$ where $x_0' = (1, x_{01}, ..., x_{0k})$.

   (3) Model specifications on sample

   With data

y	x_1	$\cdots$	x_k
y_1	x_{11}	$\cdots$	x_{1k}
$\vdots$	$\vdots$		$\vdots$
y_n	x_{n1}	$\cdots$	x_{nk}

   , let $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \in R^{n \times p}$

   where $p = k + 1$, $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} \in R^p$, $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \sim N(0, \sigma^2 I_n)$. Then

   $$y = X\beta + \epsilon \sim N(X\beta, \sigma^2 I_n).$$

2. Multiple linear regression models without intercepts

   (1) Model

   $$y = \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon, \; \epsilon \sim N(0, \sigma^2)$$

   is a multiple linear regression model without intercept.

   (2) Parameters
   $\beta \in R^p$ with $p = k$, $\sigma^2 > 0$ and
   $E(y(x_0)) = \beta_1 x_{01} + \cdots + \beta_k x_{0k} = x_0'\beta$ where $x_0' = (x_{01}, ..., x_{0k})$.

   (3) Model specifications on sample

   With data

y	x_1	$\cdots$	x_k
y_1	x_{11}	$\cdots$	x_{1k}
$\vdots$	$\vdots$		$\vdots$
y_n	x_{n1}	$\cdots$	x_{nk}

   , let $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix} \in R^{n \times p}$ where

   $p = k$, $\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$, $\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \sim N(0, \sigma^2 I_n)$. Then

   $$y = X\beta + \epsilon \sim N(X\beta, \sigma^2 I_n).$$

3. Point estimators

(1) Least square estimator $\widehat{\beta}$

$\widehat{\beta}$ is a least square estimator (LSE) for β if $Q(\beta) = \|y - X\beta\|^2 \geq \|y - X\widehat{\beta}\|^2$ for all β. Then LSE for β is $\widehat{\beta} = (X'X)^{-1}X'y$.

Proof. Let $H = X(X'X)^{-1}X'$. Then

$$\begin{aligned}\|y - X\beta\|^2 &= \|y - Hy + Hy - X\beta\|^2 = \|y - Hy\|^2 + \|Hy - X\beta\|^2 \\ &\geq \|y - Hy\|^2 = \|y - X(X'X)^{-1}X'y\|^2.\end{aligned}$$

So $\widehat{\beta} = (X'X)^{-1}X'y$.

(2) Residuals, SSE and $\widehat{y}(x_0)$

Let $\widehat{y}_i$ be the estimated $E(y_i)$. Then $\widehat{y} = \begin{pmatrix}\widehat{y}_1 \\ \vdots \\ \widehat{y}_n\end{pmatrix} = X\widehat{\beta} = Hy$.

Call $e_i = y_i - \widehat{y}_i$, $i = 1, ..., n$, the residuals. Then $e = \begin{pmatrix}e_1 \\ \vdots \\ e_n\end{pmatrix} = \begin{pmatrix}y_1 - \widehat{y}_1 \\ \vdots \\ y_n - \widehat{y}_n\end{pmatrix} = (I - H)y$.

$Q(\widehat{\beta}) = \|y - X\widehat{\beta}\|^2 = \|y - \widehat{y}\|^2 = \|e\|^2 = \sum_i e_i^2 = y'(I - H)y$ is SSE.

For the model with intercept, $E[y(x_0)] = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} = x_0'\beta$ is estimated by $\widehat{y}(x_0) = \widehat{\beta}_0 + \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_k x_{0k} = x_0'\widehat{\beta}$.

For the model without intercept, $E[y(x_0)] = \beta_1 x_{01} + \cdots + \beta_k x_{0k} = x_0'\beta$ is estimated by $\widehat{y}(x_0) = \widehat{\beta}_1 x_{01} + \cdots + \widehat{\beta}_k x_{0k} = x_0'\widehat{\beta}$.

(3) MLEs

By $y \sim N(X\beta, \sigma^2 I)$, the likelihood function of β and σ^2 is

$$L(\beta, \sigma^2) = \frac{1}{(2\pi)^{n/2}|\sigma^2 I_n|^{1/2}} \exp\left(-\frac{1}{2\sigma^2}\|y - X\beta\|^2\right).$$

$\widetilde{\beta}$ and $\widetilde{\sigma}^2$ are MLEs for β and σ^2 if

$$L(\beta, \sigma^2) \leq L(\widetilde{\beta}, \widetilde{\sigma}^2) \text{ for all } \beta \text{ and } \sigma^2.$$

Then the MLE for β is $\widetilde{\beta} = \widehat{\beta} = (X'X)^{-1}X'y$, the MLE for σ^2 is $\widetilde{\sigma}^2 = \frac{SSE}{n}$, and $L\left(\widetilde{\beta}, \widetilde{\sigma}^2\right) = \left(\frac{n}{2\pi e}\right)^{n/2} SSE^{-n/2}$.

Ex: For $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon$

i	y_i	Estimated $E(y_i) = \widehat{y}_i$	Residual $e_i = y_i - \widehat{y}_i$	e_i^2
1	y_1	$\widehat{y}_1$	$e_1 = y_1 - \widehat{y}_1$	$e_1^2 = (y_1 - \widehat{y}_1)^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	$\widehat{y}_n$	$e_n = y_n - \widehat{y}_n$	$e_n^2 = (y_n - \widehat{y}_n)^2$
Total:	c	c	0	SSE

$\sum_i y_i - \sum_i \widehat{y}_i = 1_n'(y - \widehat{y}) = (Xe_1)'(I - H)y = e_1'(X' - X'H)y = e_1'(X' - X')y = 0$.

So $\sum_i y_i = \sum_i \widehat{y}_i$, $\sum_i e_i = 0$, $\sum_i e_i^2 = \|y - X\widehat{\beta}\|^2 = Q(\widehat{\beta}) = SSE$.

L09 Unbiased estimators and t-distributions

1. Unbiased estimators (UEs)

   (1) MLE and LSE $\widehat{\beta}$ is an UE for β.

   Proof. Note that $y \sim N(X\beta,\, \sigma^2 I_n)$ and $\widehat{\beta} = Ay$ where $A = (X'X)^{-1}X'$. Then
   $\widehat{\beta} = Ay \sim N\left(AX\beta,\, A\sigma^2 I_n A'\right) = N\left(\beta,\, \sigma^2 (X'X)^{-1}\right)$ is an UE for β.

   (2) $MSE = \frac{SSE}{n-p}$ is an UE for σ^2.

   Proof. Note that $y \sim N(X\beta,\, \sigma^2 I_n)$ and $\frac{SSE}{\sigma^2} = y'Ay$ where $A = \frac{I-H}{\sigma^2}$.
   But $A\sigma^2 I_n A = A$, $(X\beta)'A(X\beta) = 0$ and $\text{tr}(A\sigma^2 I_n) = n - p$. So $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$.
   Thus $E(MSE) = E\left(\frac{\sigma^2}{n-p}\frac{SSE}{\sigma^2}\right) = \frac{\sigma^2}{n-p}E(\chi^2(n-p)) = \sigma^2$.
   Hence MSE is an UE for σ^2.

   (3) $\widehat{y}(x_0)$ is an UE for $x_0'\beta = E(y(x_0))$

   Proof. Note that $\widehat{y}(x_0) = x_0'\widehat{\beta}$ and $\widehat{\beta} \sim N\left(\beta,\, \sigma^2 (X'X)^{-1}x_0\right)$.
   Thus $\widehat{y}(x_0) \sim N\left(x_0^{\beta},\, \sigma^2 x_0'(X'X)^{-1}x_0\right)$ is an UE for $x_0'\beta = E(y(x_0))$.

   (4) $\text{Cov}(\widehat{\beta}_i) = \sigma^2\left[(X'X)^{-1}\right]_{(i,i)}$ has UE $S^2_{\widehat{\beta}_i} = \text{MSE}\ \left[(X'X)^{-1}\right]_{(i,i)}$. $MSE\,(X'X)^{-1}$.
   $S_{\widehat{\beta}_i}$ is called the standard error for $\widehat{\beta}_i$.

   (5) $\text{Cov}(\widehat{y}(x_0)) = \sigma^2 x_0'(X'X)^{-1}x_0$ has UE $S^2_{\widehat{y}(x_0)} = \text{MSE}\ x_0'(X'X)^{-1}x_0$.
   $S_{\widehat{y}(x_0)}$ is called the standard error for $\widehat{y}(x_0)$.

2. t-distributions

   (1) $\widehat{\beta}$ and SSE are independent.

   Proof. Note that $y \sim N(X\beta,\, \sigma^2 I_n)$, $\widehat{\beta} = Ay$ where $A = (X'X)^{-1}X'$,
   and $SSE = y'(I-H)y$.
   But $A\sigma^2 I_n(I-H) = 0$. So $\widehat{\beta}$ and SSE are independent.

   (2) $\frac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}_i}} \sim t(n-p)$.

   Proof. $\widehat{\beta}_i \sim N(\beta_i,\, \sigma^2[(X'X)^{-1}]_{(i,i)}) \Longrightarrow \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2\,[(X'X)^{-1}]_{(i,i)}}} \sim N(0,\, 1^2)$
   is independent to $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$. Thus
   $\frac{\widehat{\beta}_i - \beta_i}{\sqrt{\sigma^2\,[(X'X)^{-1}]_{(i,i)}}}\frac{1}{\sqrt{\frac{SSE}{\sigma^2}/(n-p)}} \sim t(n-p)$, i.e., $\frac{\widehat{\beta}_i - \beta_i}{S_{\widehat{\beta}_i}} \sim t(n-p)$.

   (3) $\frac{\widehat{y}(x_0) - x_0'\beta}{S_{\widehat{y}(x_0)}} \sim t(n-p)$

   Proof. $\widehat{y}(x_0) \sim N\left(x_0'\beta,\, \sigma^2 x_0'(X'X)^{-1}x_0\right) \Longrightarrow \frac{\widehat{y}(x_0) - x_0'\beta}{\sqrt{\sigma^2\, x_0'(X'X)^{-1}x_0}} \sim N(0,\, 1^2)$
   is independent to $\frac{SSE}{\sigma^2} \sim \chi^2(n-p)$.
   Thus $\frac{\widehat{y}(x_0) - x_0'\beta}{\sqrt{\sigma^2\, x_0'(X'X)^{-1}x_0}}\frac{1}{\sqrt{\frac{SSE}{\sigma^2}/(n-p)}} \sim t(n-p)$, i.e., $\frac{\widehat{y}(x_0) - x_0'\beta}{S_{\widehat{y}(x_0)}} \sim t(n-p)$.

3. SAS

(1) Standard output

For $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ and $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$

`proc reg;`
`model y=x1 x2;`
`run;`
	`proc reg;`
	`model y=x1 x2/noint;`
	`run;`

give standard output. In Parameter table $\widehat{\beta}_i$, $S_{\widehat{\beta}_i}$, $t = \frac{\widehat{\beta}_i}{S_{\widehat{\beta}_i}}$ and $2P(t(n-p) > |t_{ob}|)$ are listed. For the model with intercept $i = 0, 1, 2$ and $p = 3$; for the model without intercept $i = 1$, 2 and $p = 2$. In ANOVA table SSE, $n - p$ and MSE are listed.

(2) Matrix $\begin{pmatrix} X'X & X'y \\ y'X & y'y \end{pmatrix}$

For $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ and $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$

`proc reg;`
`model y=x1 x2/xpx;`
`run;`
	`proc reg;`
	`model y=x1 x2/noint xpx;`
	`run;`

display matrix $\begin{pmatrix} X'X & X'y \\ y'X & y'y \end{pmatrix} \in R^{(p+1)\times(p+1)}$. For the model with intercept this matrix is 4×4; for the model without intercept this matrix is 3×3.

(3) Matrix $\begin{pmatrix} (X'X)^{-1} & \widehat{\beta} \\ \widehat{\beta}' & \text{SSE} \end{pmatrix}$

For $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ and $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$

`proc reg;`
`model y=x1 x2/i;`
`run;`
	`proc reg;`
	`model y=x1 x2/noint i;`
	`run;`

display matrix $\begin{pmatrix} (X'X)^{-1} & \widehat{\beta} \\ \widehat{\beta}' & \text{SSE} \end{pmatrix} \in R^{(p+1)\times(p+1)}$. For the model with intercept this matrix is 4×4; for the model without intercept this matrix is 3×3.

Ex: The values of variables y, x_1 and x_2 in Table3.2 on p76 are stored in file Table32.txt. Consider model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, $\epsilon \sim N(0, \sigma^2)$.

(i) Find the value of $\widehat{\beta}_1$, the LSE and MLE for β_1.
$\widehat{\beta}_1 = (0, 1, 0)(X'X)^{-1}X'y = 1.61591$.

(ii) Find $\text{var}(\widehat{\beta}_1)$.

$$\text{var}(\widehat{\beta}_1) = [(X'X)^{-1}]_{(2,2)}\sigma^2 = 0.00274\ \sigma^2.$$

(iii) Find standard error for $\widehat{\beta}_1$.

$$S_{\widehat{\beta}_1} = \sqrt{MSE\ [(X'X)^{-1}]_{(2,2)}} = 0.17073.$$

```
SAS:      proc reg;
          model y=x1 x2/i;
          run;
```