L22: Multivariate multiple linear regression model

1. Multivariate multiple liner regression model and samples

   (1) Model

   Let $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$, $\beta = \begin{pmatrix} \beta_{10} & \beta_{11} & \cdots & \beta_{1,q-1} \\ \beta_{20} & \beta_{21} & \cdots & \beta_{2,\,q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p0} & \beta_{p1} & \cdots & \beta_{p,q-1} \end{pmatrix}$, $x = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_{q-1} \end{pmatrix}$ and $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix} \sim N(0,\, \Sigma)$. Then

   $$y = \beta x + \epsilon \text{ is a multivariate multiple linear regression model.}$$

   It is multivariate because the dependent response $y \in R^p$ with $p > 1$; it is multiple because the independent predictor $x \in R^q$ with $q > 2$; it is linear because the regression function $E(y) = \beta x$ is a linear function of unknown parameter matrix $\beta \in R^{p\times q}$.

   (2) p univariate multiple linear regression models

   The multivariate multiple linear regression implies p univariate multiple linear regressions,

   $$y_i = \beta_{i0} + \beta_{i1}x_1 + \cdots + \beta_{i,q-1}x_{q-1} + \epsilon_i \text{ with } \epsilon_i \sim N(0,\, \sigma_i^2)$$

   for $i = 1, ..., p$. These p univariate models share the same predictor vector x.

   The p univariate models do not imply the original multivariate model since the specifications of $\text{cov}(y_i,\, y_j) = \sigma_{ij}$ for $i \neq j$ in the original model are not specified by the group of p univariate models.

   (3) Samples

   Let the columns of $X \in R^{q\times n}$ be n observed values of the predictor vector $x \in R^q$, and the columns of $Y \in R^{p\times n}$ be n corresponding observed response $y \in R^p$. Then $\text{vec}(Y) \sim N(\text{vec}(\beta X),\, I_n \otimes \Sigma)$. Thus

   $$Y \sim N_{p\times n}(\beta X,\, \Sigma,\, I_n) \text{ represents data from } y = \beta x + \epsilon.$$

   (4) Samples from the univariate models

   The elements of the ith row of Y, $(y_{i1}, ..., y_{in})$ are the observed $y_i = \beta_{i0}+\beta_{i1}x_1+\cdots+\beta_{iq-1}x_{q-1}+\epsilon_i$ when x assume the values of the columns of X. Thus

   $$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix} \sim N\left(X' \begin{pmatrix} \beta_{i0} \\ \vdots \\ \beta_{iq-1} \end{pmatrix},\, \sigma_i^2 I_n \right)$$

   represents data from $y_i = \beta_{i0} + \beta_{i1}x_1 + \cdots + \beta_{iq-1}x_{q-1} + \epsilon_i$, $i = 1, ..., p$.

2. Least square estimator for parameter matrix $\beta \in R^{p\times q}$

   (1) Definition of LSE for β

   Based on $Y \sim N_{p\times n}(\beta X,\, \Sigma,\, I_n)$, $E(Y) = \beta x$. If

   $$Q(\beta) = \|Y - E(Y)\|^2 = \|Y - \beta X\|^2 = \text{tr}[(Y - \beta X)(Y - \beta X)'] \geq Q(\widehat{\beta}) \text{ for all } \beta,$$

   then $\widehat{\beta}$ is called a least square estimator (LSE) for β.

   (2) Definition of LSE for the ith row of β.

   Based on $\begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix} \sim N\left(X' \begin{pmatrix} \beta_{i0} \\ \vdots \\ \beta_{iq-1} \end{pmatrix},\, \sigma_i^2 I_n \right)$, if $\left\| \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix} - X' \begin{pmatrix} \beta_{i0} \\ \vdots \\ \beta_{iq-1} \end{pmatrix} \right\|^2 \geq \left\| \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix} - X' \begin{pmatrix} \widehat{\beta}_{i0} \\ \vdots \\ \widehat{\beta}_{iq-1} \end{pmatrix} \right\|^2$

for all $\begin{pmatrix} \beta_{i0} \\ \vdots \\ \beta_{iq-1} \end{pmatrix}$, then $\begin{pmatrix} \widehat{\beta}_{i0} \\ \vdots \\ \widehat{\beta}_{iq-1} \end{pmatrix}$ is the LSE for $\begin{pmatrix} \beta_{i0} \\ \vdots \\ \beta_{iq-1} \end{pmatrix}$, $i = 1, ..., p$.

Based on the study on univariate regression, it has been known that the LSE of the ith row of β is $\begin{pmatrix} \widehat{\beta}_{i0} \\ \vdots \\ \widehat{\beta}_{iq-1} \end{pmatrix} = (XX')^{-1}X \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix}$. Thus $(\widehat{\beta}_{i0}, ..., \widehat{\beta}_{iq-1}) = (y_{i1}, ..., y_{in})X'(XX')^{-1}$.

(3) Formula for LSE of β

Let $\widehat{\beta} = \left(\widehat{\beta}_{ij}\right)_{p\times q}$ with the ith row $(\widehat{\beta}_{i0}, ..., \widehat{\beta}_{iq-1}) = (y_{i1}, ..., y_{in})X'(XX')^{-1}$. Then $\widehat{\beta} = YX'(XX')^{-1}$. This $\widehat{\beta}$ is LSE for β.

Proof.
$$\begin{aligned} Q(\beta) &= \operatorname{tr}[(Y-\beta X)(Y-\beta X)'] = \sum_{i=1}^p \left\| \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix} - X' \begin{pmatrix} \beta_{i0} \\ \vdots \\ \beta_{iq-1} \end{pmatrix} \right\|^2 \\ &\geq \sum_{i=1}^p \left\| \begin{pmatrix} y_{i1} \\ \vdots \\ y_{in} \end{pmatrix} - X' \begin{pmatrix} \widehat{\beta}_{i0} \\ \vdots \\ \widehat{\beta}_{iq-1} \end{pmatrix} \right\|^2 = \operatorname{tr}[(Y-\widehat{\beta}X)(Y-\widehat{\beta}X)'] = Q(\widehat{\beta}). \end{aligned}$$

Comment: To get LSE of β, get LSE for each row of β in the univariate regression.

Ex: Consider model $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \beta_{10} & \beta_{11} & \beta_{12} \\ \beta_{20} & \beta_{21} & \beta_{23} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$. In the output of SAS code

```
data a;
  infile "C:\data.txt";
  input y1 y2 x1 x2;
  run;
```

```
proc reg;
  model y1 y2=x1 x2;
  run;
```

there are

For y1		For y2	
parameter	value	parameter	value
intercept	1.111	intercept	-1.111
x1	2.222	x1	-2.222
x2	3.333	x2	-3.333

Thus $\widehat{\beta} = \begin{pmatrix} 1.111 & 2.222 & 3.333 \\ -1.111 & -2.222 & -3.333 \end{pmatrix}$

3. LSE of β related statistics

(1) Estimated regression function

The regression function $E[y(x)] = \beta x$ is estimated by $\widehat{y}(x) = \widehat{\beta}x$ also called the prediction equation. If x is given, $\widehat{y}(x)$ gives the estimated mean of y.

(2) Fitted value matrix and residual matrix

With data $E(Y) = \beta X$ is estimated by the fitted value matrix $\widehat{Y} = \widehat{\beta}X = YX'(XX')^{-1}X$. Here $X'(XX')X = X^+X = H$ is called the hat matrix. $Y - \widehat{Y} = Y[I - X'(XX')^{-1}X]$ is the residual matrix.

(3) Error matrix E

$E = (Y - \widehat{Y})(Y - \widehat{Y})' = Y[I - X'(XX')^{-1}X]Y'$ is the error matrix.

L23 UEs and MLEs

1. Sampling distributions

   (1) Normal distributions

   (i) $\widehat{\beta} \sim N_{p\times q}(\beta,\, \Sigma,\, (XX')^{-1})$ (ii) $\widehat{y}(x) \sim N(\beta x,\, x'(XX')^{-1}x\,\Sigma)$

   (iii) $\widehat{Y} \sim N_{p\times n}(\beta X,\, \Sigma,\, X'(XX')^{-1}X)$ (iv) $Y - \widehat{Y} \sim N_{p\times n}(0,\, \Sigma,\, I - X'(XX')^{-1}X)$.

   Proof. Tool: $X \sim N_{p\times n}(M,\, \Sigma,\, \Psi) \Longrightarrow AXB + C \sim N_{q\times m}(AMB + C,\, A\Sigma A',\, B'\Psi B)$.

   Note that $Y \sim N_{p\times n}(\beta X,\, \Sigma,\, I_n)$.

   (i) $\widehat{\beta} = YX'(XX')^{-1} \sim N_{p\times q}(\beta,\, \Sigma,\, (XX')^{-1})$

   (ii) $\widehat{y}(x) = \widehat{\beta}x = N_{p\times 1}(\beta x,\, \Sigma,\, x'(XX')^{-1}x) = N(\beta x,\, x'(XX')^{-1}x\Sigma)$.

   (iii) $\widehat{Y} = \widehat{\beta}X \sim N_{p\times n}(\beta X,\, \Sigma,\, X'(XX')^{-1}X)$.

   (iv) $Y - \widehat{Y} = Y(I - H) \sim N_{p\times n}(\beta X(I - H),\, \Sigma,\, I - H) = N_{p\times n}(0,\, \Sigma, I - X'(XX')^{-1}X)$.

   (2) Wishart distribution: $E \sim W_{p\times p}(n - q,\, \Sigma)$.

   Proof. Tool: $X \sim N_{p\times n}(M,\, \Sigma,\, I)$, $A^2 = A = A' \Longrightarrow XAX' \sim W_{p\times p}(MAM',\, \mathrm{tr}(A),\, \Sigma)$.

   Note that $Y \sim N_{p\times n}(\beta X,\, \Sigma,\, I)$ and $(I - H)^2 = I - H = (I - H)'$. So

   $$\begin{aligned} E = Y(I-H)Y' &\sim W_{p\times p}((\beta X)(I-H)(\beta X)',\, \mathrm{tr}(I-H),\, \Sigma) = W_{p\times p}(0,\, n-q,\, \Sigma) \\ &= W_{p\times p}(n-q,\, \Sigma). \end{aligned}$$

   (3) $\widehat{\beta}$ and E are independent.

   Proof. Tool: Under $X \sim N_{p\times n}(M,\, \Sigma,\, \Psi)$,

   $$A_1XB_1 \text{ and } A_2XB_2 \text{ are independent} \Longleftrightarrow A_1\Sigma A_2' = 0 \text{ or } B_1'\Psi B_2 = 0.$$

   So $\widehat{\beta} = YX'(XX')^{-1}$ and $Y - \widehat{Y} = Y(I - H)$ are independent since $Y \sim N_{p\times n}(\beta X,\, \Sigma,\, I)$ and $[X'(XX')^{-1}]'I_n(I - H) = 0$. Consequently $\widehat{\beta}$ and $E = Y(I - H)[Y(I - H)]'$ are independent.

   Ex1: Suppose $X \in R^{q\times n}$ has full row rank q. Then

   LSE of β, $\widehat{\beta}$, is an UE for β since $E(\widehat{\beta}) = E[N_{p\times q}(\beta,\, \Sigma,\, (XX')^{-1})] = \beta$.a

   $\frac{E}{n-q}$ is an UE for Σ since $E\left(\frac{E}{n-q}\right) = \frac{1}{n-q}E(E) = \frac{1}{n-q}E[W_{p\times p}(n - q,\, \Sigma)] = \frac{1}{n-q}\,(n - q)\Sigma = \Sigma$.

2. MLEs of β and Σ

   (1) Maximizing the likelihood function: Step I

   With LSE $\widehat{\beta} = YX'(XX')^{-1}$ and $E = (Y - \widehat{\beta}X)(Y - \widehat{\beta}X)'$, from $Y \sim N_{p\times n}(\beta X,\, \Sigma,\, I_n)$,

   $$\begin{aligned} L(\beta,\, \Sigma) &= \tfrac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\left[(Y - \beta X)'\Sigma^{-1}(Y - \beta X)'\right]\right\} \\ &= \tfrac{|\Sigma^{-1}|^{n/2}}{(2\pi)^{np/2}} \exp\left\{-\tfrac{1}{2}\mathrm{tr}\left[\Sigma^{-1/2}(Y - \beta X)(Y - \beta X)'\Sigma^{-1/2}\right]\right\}. \end{aligned}$$

   But $(Y - \beta X)(Y - \beta X)' = E + (\widehat{\beta}X - \beta X)(\widehat{\beta}X - \beta X)$. So

   $$\begin{aligned} L(\beta,\, \Sigma) &= \tfrac{|\Sigma^{-1}|^{n/2}}{(2\pi)^{np/2}} \exp\left[-\tfrac{1}{2}\mathrm{tr}(\Sigma^{-1/2}E\Sigma^{-1/2})\right] \cdot \exp\left\{-\tfrac{1}{2}\mathrm{tr}\left[\Sigma^{-1/2}(\widehat{\beta}X - \beta X)(\widehat{\beta}X - \beta X)'\Sigma^{-1/2}\right]\right\} \\ &\le \tfrac{|\Sigma^{-1/2}E\Sigma^{-1/2}|^{n/2}}{(2\pi)^{np/2}|E|^{n/2}} \exp\left[-\tfrac{1}{2}\mathrm{tr}(\Sigma^{-1/2}E\Sigma^{-1/2})\right] = L(\widehat{\beta},\, \Sigma). \end{aligned}$$

   (2) Maximizing the likelihood function: Step II

   Let $\Sigma^{-1/2}E\Sigma^{-1/2} = P\Lambda P'$ be the EVD. Then

   $$\begin{aligned} L(\widehat{\beta},\, \Sigma) &= \tfrac{(\lambda_1\cdots\lambda_p)^{n/2}}{(2\pi)^{np/2}|E|^{n/2}} \exp\left(\tfrac{\lambda_1 + \cdots + \lambda_p}{2}\right) \\ &= \tfrac{1}{(2\pi)^{np/2}|E|^{n/2}} \prod_{i=1}^{p} f(\lambda_i) \end{aligned}$$

   where $f(\lambda_i) = \lambda_i^{n/2}e^{-\frac{\lambda_i}{2}}$ is maximized when $\lambda_i = n$, i.e.,

   $\Sigma^{-1/2}E\Sigma^{-1/2} = PnI_nP' = nI_n \Longleftrightarrow \Sigma = \frac{E}{n}$. Hence

   $$L(\widehat{\beta},\, \Sigma) \le L\left(\widehat{\beta},\, \frac{E}{n}\right) = \left(\frac{n}{2\pi e}\right)^{np/2} |E|^{-n/2}$$

(3) Conclusions

$\widehat{\beta}$ is MLE for β, $\frac{E}{n}$ is MLE for Σ, and $\max[L(\beta, \Sigma) : \beta, \Sigma] = \left(\frac{n}{2\pi e}\right)^{np/2} \cdot |E|^{n/2}$.

3. Computations

Consider $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \beta_{10} & \beta_{11} & \beta_{12} \\ \beta_{20} & \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$ with data

y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3
x_1	-2	-1	0	1	2
x_2	3	4	2	1	5

We need to find $\widehat{\beta}$, $\widehat{Y}$ and $Y - \widehat{Y}$. Based on the result we can calculate E.

(1) Enter data

```
data a;
  infile "D:\\Example.txt"
  input y1 y2 x1 x2;
```

Data file:

```
5 -3 -2 3
3 -1 -1 4
4 -1 0 2
2 2 1 1
1 3 2 5
```

(2) Find $\widehat{\beta}$

```
proc reg;
  model y1 y2=x1 x2;
  run;
```

In the parameter table for the univariate model for y_1, the first row of $\widehat{\beta}$, $\widehat{\beta}_{10}$, $\widehat{\beta}_{11}$ and $\widehat{\beta}_{12}$ are displayed. In the output for y_2, the second row of $\widehat{\beta}$ is displayed.

(3) Find $\widehat{Y}$ and $Y - \widehat{Y}$.

```
proc reg;
  model y1 y2=x1 x2/p;
  run;
```

The first row of $\widehat{Y}$ and the first row of $Y - \widehat{Y}$ are in the output for the model for y_1. The second rows of $\widehat{Y}$ and $Y - \widehat{Y}$ are in the output for the model for y_2.

(4) Find $\widehat{y}(3, 2)$

$\widehat{y}(3, 2) = \widehat{\beta} \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} = \widehat{\beta}_0 + \widehat{\beta}_1 \cdot 3 + \widehat{\beta}_2 \cdot 2 \in R^2$ is the value of estimated regression function when $x_1 = 3$ and $x_2 = 2$. It is the estimated $E[y(3, 2)]$.

```
data b;
  input y1 y2 x1 x2;
  datalines;
  . . 3 2
  ;
```

```
data c;
   set a b;
proc reg;
   model y1 y2=x1 x2/p;
   run;
```

The first component of $\widehat{y}(3, 2)$ is in the output for model for y_1. The second component of $\widehat{y}(3, 2)$ is in that for y_2.

Comments: In the output for (4) one can find $\widehat{\beta}$, $\widehat{Y}$, $Y - \widehat{Y}$ and $\widehat{y}(3, 2)$. While $E = (Y - \widehat{Y})(Y - \widehat{Y})'$ can be calculated from $Y - \widehat{Y}$, we will see how to ask SAS to display E.