L16: Prediction function

1. Prediction function

   (1) Prediction function
   For $\mathbf{y}(\mathbf{x}) = B'\mathbf{x}+\epsilon$, with B estimated by $\widehat{B} = (X'X)^{-1}X'Y$ based on data matrices $Y \in R^{n\times p}$ and $X \in R^{n\times q}$, the regression function $E(\mathbf{y}(\mathbf{x})) = B'\mathbf{x}$ is estimated by prediction function $\widehat{y}(\mathbf{x}) = \widehat{B}'\mathbf{x}$. So when $\mathbf{x}$ is given, the estimated $E(y)$ and predicted $y(x)$, $\widehat{y}(x)$, can be calculated.

   (2) Fitted value matrix
   With available $\mathbf{x}_i \in R^q$, $i = 1, ..., n$, in $X' = (\mathbf{x}_1, .., \mathbf{x}_n) \in R^{q\times n}$,

   $$(\widehat{y}(\mathbf{x}_1), ..., \widehat{y}(\mathbf{x}_n)) = \widehat{B}'X' \overset{Denoted}{=\!=} \widehat{Y}'.$$

   Here $\widehat{Y} = X\widehat{B} = X(X'X)^{-1}X'Y$ is the fitted value matrix. This matrix is estimated $E(Y) = XB$.

   (3) Residual matrix
   In $Y = XB+\mathcal{E}'$, $\mathcal{E}' = Y - XB$ is predicted by $Y - X\widehat{B} = Y - \widehat{Y} = [I - X(X'X)^{-1}X']Y$ called the residual matrix. In $Y'[I - X(X'X)^{-1}X'] = (\mathbf{y}_1 - \widehat{y}(\mathbf{x}_1), ..., \mathbf{y}_n - \widehat{y}(\mathbf{x}_n))$, the columns are residuals based on the data from which $\widehat{B}$ was obtained.

   (4) Error matrix
   Recall that the error matrix $E = (Y - X\widehat{B})'(Y - X\widehat{B}) = (Y - \widehat{Y})'(Y - \widehat{Y}) = Y'[I - X(X'X)^{-1}X']Y$. MLE of Σ is $\widehat{\Sigma} = \frac{E}{n}$ and the UE for Σ is $\frac{E}{n-p}$.

2. Computation for $\widehat{Y}$ and $Y - \widehat{Y}$.

   (1) Computing $\widehat{Y}$ and $Y - \widehat{Y}$ column after column.
   For $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = B'\mathbf{x} + \epsilon$,
   the ith column of $\widehat{Y}$ is produced by the univariate regression of the ith component of $\mathbf{y}$ on $\mathbf{x} \in R^q$,

   $$\widehat{Y} = X\widehat{B} = X(X'X)^{-1}X'Y \Longleftrightarrow \widehat{Y}e_i = X(X'X)^{-1}X'Ye_i, \ i = 1, .., p.$$

   The ith column of $Y - \widehat{Y}$ is produced by the univariate regression of the ith component of $\mathbf{y}$ on $\mathbf{x} \in R^q$,

   $$Y - \widehat{Y} = [I - X(X'X)^{-1}X']Y \Longleftrightarrow (Y - \widehat{Y})e_i = [I - X(X'X)^{-1}X']Ye_i, \ = 1, .., p.$$

   (2) SAS

```
proc reg;
   model y1 y2 y3=x1 x2/p;
   run;
```

```
proc reg;
  model y1 y2 y3=x1 x2/noint p;
  run;
```

   Here "p" stands for Predict. In the output, the three columns of Y and $Y - \widehat{Y}$ will be displayed in the output for the univariate regression for y_i, $i = 1, 2, 3$.

   Comment: Once $Y - \widehat{Y}$ is available, $E = (Y - \widehat{Y})'(Y - \widehat{Y})$. But since $Y - \widehat{Y} \in R^{p\times n}$, with large n this method may not be practical. We will see later that SAS can display E for us.

(3) Use prediction function

With given new values for $\mathbf{x} = \mathbf{x}_p$, the $\mathbf{y}(\mathbf{x}_p)$ and $E[\mathbf{y}(\mathbf{x}_p)]$ are predicted and estimated by the same value of the prediction function $\widehat{\mathbf{y}}(\mathbf{x}_p) = \widehat{B}'\mathbf{x}_p$. SAS can calculate this value component after component in univariate regression for each component of $\mathbf{y}$.

```
data a;
  infile "D:\Ex.txt";
  input y1 y2 x1 x2 @@;
data b;
  input y1 y2 x1 x2;
  datalines;
  . . 1.2 -3.5
  . . 2   -6
  ;
```

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

3. Hypothesis testing

(1) Matrix B

In $\mathbf{y} = B'\mathbf{x} + \epsilon$, each component of $B \in R^{q\times p}$ is associated with a particular predictor in $\mathbf{x} \in R^q$ and a particular component of $\mathbf{y} \in R^p$.

Ex1: For $\mathbf{y} \in R^2$, in $\mathbf{y} = B' \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix} + \epsilon$, $\begin{array}{cc} & \begin{array}{cc} y1 & y2 \end{array} \\ \begin{array}{c} 1 \\ x1 \\ x2 \end{array} & \begin{bmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \end{array} = B.$

(2) Hypotheses

Many hypothesis are rendered as $H_1 B H_2 = D$.

Ex2: For the model in Ex1,

(i) x_1 does not contribute anything to $y_2 \iff \beta_{12} = 0 \iff (0,\ 1,\ 0)B \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$.

(ii) When x_1 increases by 1, the increment in $E(y_1)$ and in $E(y_2)$ are equal $\iff \beta_{11} = \beta_{12}$ $\iff (0,\ 1,\ 0)B \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0$.

(iii) x_1 contributes nothing to the model $\iff \beta_{11} = \beta_{12} = 0 \iff (0,\ 1,\ 0)BI_2 = (0,\ 0)$.

(iv) The model is useless $\iff \beta_{ij} = 0$ for all $i = 1, 2$ and $j = 1, 2 \iff (0,\ I_2)B = \mathbf{0}$.

(3) Likelihood ratio and Wilk's Lambda

From sample $Y = XB + \mathcal{E}'$ where $\mathcal{E} \sim N_{n\times p}(XB,\ I_n,\ \Sigma)$, the likelihood function $L(B,\ \Sigma) \le L(\widehat{B},\ \widehat{\Sigma}) = \left(\frac{n}{2\pi e}\right)^{np/2} |E|^{-n/2}$ where the error matrix $E = (Y - X\widehat{B})'(Y - X\widehat{B})$.

Under H_0, the model is converted to a reduced model with error matrix $E_r = (Y - X\widehat{B}_0)'(Y - X\widehat{B}_0)$ and $\max[L(B,\ \Sigma) :\ H_0] = \left(\frac{n}{2\pi e}\right)^{np/2} |E_r|^{-n/2}$.

Thus LR $= \frac{\max[L(B,\,\Sigma):]}{\max[L(B,\,\Sigma):\,H_0]} = \left(\frac{|E|}{|E_r|}\right)^{-n/2}$ is a decreasing function of Wilk's Lambda, $\Lambda = \frac{|E|}{|E_r|}$.

(4) LRT scheme

$H_0 :,\ H_1 B H_2 = D$ versus $H_a :\ H_1 B H_2 \ne D$

Test statistic: $\Lambda = \frac{|E|}{|E_r|}$

p-value: $P(\Lambda \le \Lambda_{ob} \mid H_0)$

L17: Tests in multivariate regression

1. SAS for multivariate tests

   (1) Test scheme

   H_0 : —- versus H_a : —-
   Test Statistic: $\Lambda = \frac{|E|}{|E_r|}$
   p-value: $P(\Lambda \le \Lambda_{ob}|H_0)$.

   p-value: $P(\Lambda \le \Lambda_{0b}|H_0) = P(\Lambda \le --|H_0) \approx P(F(--,--) > --) = --$
   Conclusion —-

   To complete the report fill in the blanks.

   (2) SAS mtest statement

   Consider model $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \end{pmatrix} + \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \end{pmatrix} x_1 + \begin{pmatrix} \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{pmatrix} x_2 + \begin{pmatrix} \beta_{31} \\ \beta_{32} \\ \beta_{33} \end{pmatrix} x_3 + \epsilon$

   For $H_0: \beta_1 = 0$ and $\beta_2 = 0$ versus $H_a: \beta_1 \ne 0$ or $\beta_2 \ne 0$

```
proc reg;
     model y1 y2 y3=x1 x2 x3/noprint;
     mtest x1, x2;
     run;
```

   For $H_0: \beta_{11} = \beta_{12} = \beta_{13}$ versus $H_a: \ \beta_{1i} \ne \beta_{1j}$ for some i, j

```
proc reg;
     model y1 y2 y3=x1 x2 x3/noprint;
     mtest y1-y2, y2-y3, x1/printe;
     run;
```

   printe: Print matrix E.
   For $H_0: \beta_{i1} = \beta_{i2}$ for all $i = 1, 2, 3$ versus $H_a: \beta_{i1} \ne \beta_{i2}$ for some $i = 1, 2, 3$

```
proc reg;
     model y1 y2 y3=x1 x2 x3/noprint;
     mtest y1-y2, x1, x2, x3;
     run;
```

   For $H_0: \beta_{21} = \beta_{22} + 1$ versus $H_a: \beta_{21} \ne \beta_{22} + 1$

```
proc reg;
     model y1 y2 y3=x1 x2 x3/noprint;
     mtest y1-y2, x2=1;
     run;
```

   (3) Presenting report

   $H_0: \beta_{11} = \beta_{12} = \beta_{13}$ vs $H_a: \beta_{1i} \ne \beta_{1j}$ for some $i, j = 1, 2, 3$
   Test Statistic: $\Lambda = \frac{|E|}{|T|}$
   p-value: $P(\Lambda < \Lambda_{ob}|H_0)$

   $\Lambda = 0.9460$
   P-value: $P(\Lambda \le 0.9460|H_0) = P(F(2, 15) > 0.43) = 0.6596$
   Fail to reject H_0
   No evidence against $\beta_{11} = \beta_{21} = \beta_{31}$

2. Other statistics produced by mtest

   (1) Matrices E_r and H
   $$\begin{aligned} E_r &= (Y - X\widehat{B}_0)'(Y - X\widehat{B}_0) = (Y - X\widehat{B} + X\widehat{B} - X\widehat{B}_0)'(Y - X\widehat{B} + X\widehat{B} - X\widehat{B}_0) \\ &= (Y - X\widehat{B})'(Y - X\widehat{B}) + (\widehat{B} - \widehat{B}_0)'X'X(\widehat{B} - \widehat{B}_0) = E + H = T \end{aligned}$$
   In $T = E_r = E + H$, H is the difference between E and E_r caused by the Hypothesis.

   (2) Other three statistics
   Three other statistics are also proposed as test statistics and displayed by SAS. They are

Statistics	value
Wilk's Lambda	$\frac{\lvert E\rvert}{\lvert E+H\rvert}$
Pillai's Trace	$\mathrm{tr}[H(E+H)^{-1}]$
Hotelling-Lawley Trance	$\mathrm{tr}(HE^{-1})$
Roy's greatest root	Largest eigenvalue of $E^{-1/2}HE^{-1/2}$

   (3) Expressions using eighenvalues
   Let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$ be the eigenvalues of $E^{-1/2}HE^{-1/2} = Q\Lambda Q'$.

   (i) Roy's greatest root: λ_1.

   (ii) Hotelling-Lawley trace: $\mathrm{tr}(HE^{-1}) = \mathrm{tr}(E^{-1/2}HE^{-1/2}) = \lambda_1 + \cdots + \lambda_p$.

   (iii) Pillai's trace: $\frac{\lambda_1}{1+\lambda_1} + \cdots + \frac{\lambda_p}{1+\lambda_p}$
   $$\begin{aligned} \mathrm{tr}[H(E+H)^{-1}] &= \mathrm{tr}[(E^{1/2}(E^{-1/2}HE^{-1/2})E^{1/2}(E+H)^{-1}] \\ &= \mathrm{tr}\{(E^{-1/2}HE^{-1/2})[E^{1/2}(E+H)^{-1}E^{1/2}]\} \\ &= \mathrm{tr}[(E^{-1/2}HE^{-1/2})(I + E^{-1/2}HE^{-1/2})^{-1}] \\ &= \mathrm{tr}[Q\Lambda Q'Q(I+\Lambda)^{-1}Q'] = \mathrm{tr}[\Lambda(I+\Lambda)^{-1}] = \textstyle\sum_i \frac{\lambda_i}{1+\lambda_i} \end{aligned}.$$

   (iv) Wilk's Lambda $\Lambda = \frac{|E|}{|E+H|} = \frac{1}{|I+E^{-1/2}HE^{-1/2}|} = \frac{1}{(1+\lambda_1)\cdots(1+\lambda_p)}$.

   (4) Different approaches
   Generally there are no function relations for the four statistics. So they represent 4 different test schemes and hence may produce different p-values. Wilk's Lambda is LRT statistic with approximated p-value produced.

3. Special cases

   (1) Cases of rank$(H) = 1$.
   rank$(H) = 1 \Longrightarrow$ rank$(E^{-1/2}HE^{-1/2}) = 1 \Longrightarrow \lambda_1 > 0$ and $\lambda_i = 0$ for $i = 2, .., p$.
   So Wilk's Lambda= $\frac{1}{1+\lambda_1}$. Pillai's trace= $\frac{\lambda_1}{1+\lambda_1} = 1 - \frac{1}{1+\lambda_1}$
   Hotelling-Lawley trace= Roy's greatest root= λ_1.
   In such a case the four statistics are functions each other. The four tests are hence equivalent.

   (2) One-sample tests
   $\mathbf{y} \sim N(\mu, \Sigma)$ can be written as regression $\mathbf{y} = \mu + \epsilon$, i.e., $B' = \mu$ and $\mathbf{x} = 1$.
   With data $Y \in R^{n\times p}$, $X = 1_n \in R^{n\times 1}$, $E = Y'\left(I - \frac{1_n 1_n'}{n}\right)Y =$ CSSCP.
   Under $H_0:\ \mu = 0$, $\mathbf{y} = \epsilon$, with data $Y \in R^{n\times p}$ without X,
   $$E_r = Y'Y = Y'\left(I - \frac{1_n 1_n'}{n} + \frac{1_n 1_n'}{n}\right)Y = E + H$$
   where $H = n\bar{\mathbf{x}}\,\bar{\mathbf{x}}'$ has rank 1. Thus four tests displayed by SAS are equivalent.

   (3) For the case in (2)
   $\Lambda = \left(1 + \frac{T^2}{n-1}\right)^{-1}$ where $T^2 = \overline{x}'\left(\frac{S_u}{n}\right)^{-1}\overline{x} \overset{H_0}{\sim} T^2(p, n-1) = \frac{(n-1)p}{n-p}F(p, n-p)$.
   Hence the p-values displayed are true without approximation.