L24 LRTs in multivariate regression

1. Model and hypotheses

   Many questions of interests in regression are related to hypotheses testing. Consider model

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \beta_{10} & \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{20} & \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{30} & \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix}.$$

   (1) Usefulness of the model

   The model is useless $\iff$ None of the regressors x1, x2, x3 has impact on $E(y)$
   $\iff$ $\beta_1 = 0$, $\beta_2 = 0$ and $\beta_3 = 0$
   $\iff$ $\beta_{ij} = 0$ for all $i = 1, 2, 3$ and all $j = 1, 2, 3$.

   (2) Response mean is 0 when $x_i = 0$ for all i

   $\iff \beta_0 = 0 \iff \beta_{i0} = 0$ for all i

   (3) The contributions of x_1 and x_2 to $E(y)$ are equal

   $\iff \beta_1 = \beta_2 \iff \beta_{i1} = \beta_{i2}$ for all i = 1, 2, 3..

   The increment in $E(y)$ caused by the increment in x_i by 1 unit can be used to measure the contribution of x_i to $E(y)$.

   (4) The contributions from x_1 to $E(y_1)$, $E(y_2)$ and $E(y_3)$ are equal

   $\iff \beta_{1i} = \beta_{1j}$ for all i, j

2. Likelihood ratio tests (LRTs)

   (1) Likelihood ratio

   With data suppose E is the error matrix in the original model and E_r is the error matrix from the reduced model. Then

   $\max[L(\beta, \Sigma) : \beta, \Sigma] = \left(\frac{n}{2\pi e}\right)^{np/2} |E|^{-n/2}$ and $\max[L(\beta, \Sigma) : \text{under } H_0] = \left(\frac{n}{2\pi e}\right)^{np/2} |E_r|^{-n/2}$.

   So the likelihood ratio

$$\text{LR} = \frac{\max[L(\beta, \Sigma) : \text{under } H_0]}{\max[L(\beta, \Sigma) : \beta, \Sigma]} = \left(\frac{|E|}{|E_r|}\right)^{n/2}$$

   is an increasing function of Wilks Lambda

$$\Lambda = \frac{|E|}{|E_r|} = \frac{|E|}{|E+H|}$$

   (2) Likelihood ratio test

H_0 versus H_a
Test Statistic: $\Lambda = \frac{\vert E\vert}{\vert E+H\vert}$
Reject H_0 if $\Lambda < c$.

   is a LRT. For α-level test, c should be determined by

$$P(\Lambda < c | H_0 \text{ is true}) \leq \alpha.$$

   (3) Tests by p-values

H_0 versus H_a
Test Statistic: $\Lambda = \frac{\vert E\vert}{\vert E+H\vert}$
p-value: $P(\Lambda \leq \Lambda_{ob} \vert H_0)$.

3. Implementation via SAS

   Consider the model in 1

   (1) SAS mtest statement.

(i) For $H_0: \beta_1 = 0$ and $\beta_2 = 0$ versus $H_a: \beta_1 \neq 0$ or $\beta_2 \neq 0$

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

So for $H_0: \beta_i = 0$ for all $i = 1, 2, 3$ one can use [mtest x1, x2, x3;]
But because it is a global test, one can even use [mtest ;]

(ii) For $H_0: \beta_{i1} = \beta_{j1}$ for all i, j versus $H_a: \beta_{i1} \neq \beta_{j1}$ for some i, j

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

(iii) For $H_0: \beta_{1j} = \beta_{2j}$ for all j versus $H_a: \beta_{1j} \neq \beta_{2j}$ for some j

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

(2) SAS output
With Wilk's Lambda= $\frac{|E|}{|E+H|}$, Pillai's trace= $\text{tr}[H(E+H)^{-1}]$, Hotelling-Lawley trace= $\text{tr}(HE^{-1})$ and Roy's greatest root= the largetst eigenvalue of $E^{-1/2}HE^{-1/2}$, SAS displays

```
Statistics               Value   F-value Num DF  Den DF   Pr>F
Wilks' Lambda           ------   -----   ----     ----    -----
Pillai's Trace          ------   -----   ----     ----    -----
Hotellig-Lawley Trace   ------   -----   ----     ----    -----
Roy's Greatest Root     ------   -----   ----     ----    -----
```

The results from the four statistics may not be the same. All results are based on approximation. We stick to the result from the test statistic Λ

(3) A special case
Based on sample $Y \sim N_{p\times n}(\mu 1_n', \Sigma, I_n)$ from $N(\mu, \Sigma)$, $E = Y'(I - 1_n 1_n^+)Y \sim W_{p\times p}(n-1, \Sigma)$ which is independent to $\widehat{\mu} = Y 1_n (1_n' 1_n)^{-1} = \overline{Y} \sim N(\mu, \frac{1}{n}\Sigma)$.
Under $H_0: \mu = 0$ from $N_{p\times n}(0, \Sigma, I_n)$, $E_r = Y'Y$. Thus $H = Y' 1_n 1_n^+ Y = n\overline{Y}\,\overline{Y}'$ where $\sqrt{n}\overline{Y} \overset{H_0}{\sim} N(0, \Sigma)$. Hence $T^2 = n\overline{Y}' (S)^{-1} \overline{Y} = \overline{Y}' \left(\frac{S}{n}\right)^{-1} \overline{Y} \overset{H_0}{\sim} T^2(p, n-1)$.
From $\begin{vmatrix} 1 & -n\overline{Y}' \\ \overline{Y} & E \end{vmatrix}$, $|E|\left(1 + n\overline{Y}' E \overline{Y}\right) = |E+H|$. So $\Lambda = \left(1 + \frac{T^2}{n-1}\right)^{-1}$.

Ex: For model in 1 with x_1, x_2, x_3, y_1, y_2, y_3 in Table 7.5 on page 424. By SAS in 3

$H_0: \beta_{i1} = \beta_{j1}$ for all $i, j = 1, 2, 3$ vs $H_a: \beta_{i1} \neq \beta_{j1}$ for some $i, j = 1, 2, 3$ Test Statistic: $\Lambda = \frac{|E|}{|E+H|}$ p-value: $P(\Lambda < \Lambda_{ob} \mid H_0)$
$\Lambda = 0.9460$ P-value: $P(\Lambda \leq 0.9460 \mid H_0) = P(F(2, 15) > 0.43) = 0.6596$ Fail to reject H_0 No evidence against $\beta_{11} = \beta_{21} = \beta_{31}$

L25: The concept of regression

1. Regression function

   (1) Regression model and regression function
   $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{q-1} x_q + \epsilon$ where $y \in R^p$, $\beta_i \in R^p$, $i = 0, ..., q-1$, $\epsilon \sim N(0, \Sigma)$ is a multivariate regression model.

   Let $\beta = (\beta_0, ..., \beta_{q-1}) \in R^{p\times q}$ and $x = \begin{pmatrix} 1 \\ x_0 \end{pmatrix} \in R^q$ where $x_0 = \begin{pmatrix} x_1 \\ \vdots \\ x_{q-1} \end{pmatrix} \in R^{q-1}$.

   Then the model can be written as $y = \beta x + \epsilon \sim N(X\beta, \Sigma)$
   $E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_{q-1} x_{q-1} = \beta x$ is a function of x_0.
   This function is called the regression function of y on x_0.

   (2) Sample and estimated β
   Let the columns of $Y \in R^{p\times n}$ be n observations on y when the observations on $\begin{pmatrix} 1 \\ x_0 \end{pmatrix}$ are the columns of $X = \begin{pmatrix} 1_n' \\ X_0 \end{pmatrix} \in R^{q\times n}$.
   Then $Y \sim N_{p\times n}(\beta X, \Sigma, I_n)$ characterizes the distribution of the samples.
   By least square method, β is estimated by its LSE

   $$\widehat{\beta} = YX^+ = YX'(XX')^{-1} = Y(1_n\, X_0') \begin{pmatrix} n & 1_n' X_0' \\ X_0 1_n & X_0 X_0' \end{pmatrix}^{-1}.$$

   (3) Estimated regression function
   Thus the estimated regression function is

   $$g(x_0) = (Y1_n, YX_0') \begin{pmatrix} n & 1_n' X_0' \\ X_0 1_n & X_0 X_0' \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ x_0 \end{pmatrix}$$

2. A different definition of regression function

   (1) A different definition of regression
   Suppose x_0 and y have a joint distribution. Then the conditional mean $E(y|x_0)$ is a function of x_0. This function is called the regression function of y on x_0.
   If $\begin{pmatrix} x_0 \\ y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{x_0} \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{x_0} & \Sigma_{x_0 y} \\ \Sigma_{y x_0} & \Sigma_y \end{pmatrix} \right)$, then

   $$y|x_0 \sim N\left(\mu_y - \Sigma_{yx_0}\Sigma_{x_0}^{-1}(x_0 - \mu_{x_0},\, \Sigma_y - \Sigma_{yx_0}\Sigma_{x_0}^{-1}\Sigma_{x_0 y}\right).$$

   So $E(y|x_0) = \mu_y + \Sigma_{yx_0}\Sigma_{x_0}^{-1}(x_0 - \mu_{x_0})$ is the regression function of y on x_0.

   (2) Samples and estimated μ_{x_0}, μ_y and Σ.
   Let the columns of $\begin{pmatrix} X_0 \\ Y \end{pmatrix} \in R^{(p+q-1)\times n}$ be the n observations on $\begin{pmatrix} x_0 \\ y \end{pmatrix}$.
   Then $\begin{pmatrix} \mu_{x_0} \\ \mu_y \end{pmatrix}$ is estimated by $\begin{pmatrix} \overline{x}_0 \\ \overline{y} \end{pmatrix} = \begin{pmatrix} X_0 1_n/n \\ Y 1_n/n \end{pmatrix}$
   Let $E = \begin{pmatrix} E_{x_0} & E_{x_0 y} \\ E_{y x_0} & E_y \end{pmatrix} = \begin{pmatrix} X_0 \\ Y \end{pmatrix} (I_n - 1_n 1_n^+) \begin{pmatrix} X_0 \\ Y \end{pmatrix}'$.
   Then $S = \begin{pmatrix} S_{x_0} & S_{x_0 y} \\ S_{y x_0} & S_y \end{pmatrix} = \frac{E}{n-1}$ is an UE for $\Sigma = \begin{pmatrix} \Sigma_{x_0} & \Sigma_{x_0 y} \\ \Sigma_{y x_0} & \Sigma_y \end{pmatrix}$ and $\widehat{\Sigma} = \frac{E}{n}$ is MLE for Σ.

   (3) Estimated regression function
   In the regression function $E(y|x_0) = \mu_y + \Sigma_{yx_0}\Sigma_{x_0}^{-1}(x_0 - \mu_{x_0})$, replace μ_{x_0} and μ_y by their estimators $\overline{x}_0$ and $\overline{y}$; and replace $\Sigma_{yx_0}\Sigma_{x_0}^{-1}$ by $S_{yx_0}S_{x_0}^{-1} = \widehat{\Sigma}_{yx_0}\widehat{\Sigma}_{x_0}^{-1} = [Y(I - 11^+)X_0'][X_0(I - 11^+)]X_0']^{-1}$, we obtain the estimated regression function

   $$\overline{y} + [Y(I - 11^+)X_0'][X_0(I - 11^+)X_0']^{-1}(x_0 - \overline{x}_0)$$

3. The identical results
   We show that the two approaches reach identical results on the estimated regression function

   (1) A tool

   For $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, let $A_{22.1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$.

   Then $|A| = |A_{11}| \cdot |A_{22.1}|$ and $A^{-1} = \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}$.

   Proof. By direct computation, $\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22.1} \end{pmatrix}$.

   Taking the determinants of the both sides of the equation leads to $|A| = |A_{11}| \cdot |A_{22.1}|$.
   Taking the inverses of the both sides of the equation and then solve the resulted equation for A^{-1} leads to $A^{-1} = \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22.1}^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix}$.

   Ex: For $A = \begin{pmatrix} n & 1_n'X_0' \\ X_0 1_n & X_0X_0' \end{pmatrix}$, $A_{11}^{-1} = \frac{1}{n}$, $A_{22.1}^{-1} = [X_0(I - 11^+)X_0]^{-1}$, $-A_{11}^{-1}A_{12} = -\frac{1_n'X_0'}{n}$ and $-A_{21}A_{11}^{-1} = -\overline{x}_0$. Thus $A^{-1} = \begin{pmatrix} 1 & \frac{1_n'X_0'}{n} \\ 0 & I \end{pmatrix} \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & [X_0(I - 11^+)X_0']^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\overline{x}_0 & I \end{pmatrix}$.

   Comment: For $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, let $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$.

   Then $|A| = |A_{11.2}| \cdot |A_{22}|$ and $A^{-1} = \begin{pmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{pmatrix} \begin{pmatrix} A_{11.2}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix}$.

   The proof follows from $\begin{pmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{pmatrix} = \begin{pmatrix} A_{11.2} & 0 \\ 0 & A_{22} \end{pmatrix}$

   (2) Theorem

$$
\begin{aligned}
g(x_0) &= (Y1_n, YX_0') \begin{pmatrix} n & 1_n'X_0' \\ X_0 1_n & X_0X_0' \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ x_0 \end{pmatrix} \\
&= \overline{y} + [Y(I_n - 1_n 1_n^+)X_0'][X_0(I_n - 1_n 1_n^+)X_0']^{-1}(x_0 - \overline{x}_0).
\end{aligned}
$$

   Proof By the result in Ex,

$$
\begin{aligned}
g(x_0) &= (Y1_n, YX_0') \begin{pmatrix} n & 1_n'X_0' \\ X_0 1_n & X_0X_0' \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ x_0 \end{pmatrix} \\
&= (Y1_n, YX_0') \begin{pmatrix} 1 & -\frac{1_n'X_0'}{n} \\ 0 & I \end{pmatrix} \begin{pmatrix} n^{-1} & 0 \\ 0 & [X_0(I_n - 1_n 1_n^+)X_0']^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\overline{x}_0 & I \end{pmatrix} \begin{pmatrix} 1 \\ x_0 \end{pmatrix} \\
&= (Y1_n, Y(I_n - 1_n 1_n^+)X_0') \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & [X_0(I_n - 1_n 1_n^+)X_0']^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ x_0 - \overline{x}_0 \end{pmatrix} \\
&= \left(\overline{y}, [Y(I_n - 1_n 1_n^+)X_0][X_0(I_n - 1_n 1_n^+)X_0']^{-1}\right) \begin{pmatrix} 1 \\ x_0 - \overline{x}_0 \end{pmatrix} \\
&= \overline{y} + [Y(I_n - 1_n 1_n^+)X_0'][X_0(I_n - 1_n 1_n^+)X_0']^{-1}(x_0 - \overline{x}_0).
\end{aligned}
$$

   (2) A note on E and H by SAS

   (i) "print" option in the following SAS will print matrices E and H

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

   (ii) $E \in R^{2\times 2}$ and $H \in R^{2\times 2}$ since the model for $\begin{pmatrix} y_1 - y_2 \\ y_3 \end{pmatrix} \in R^2$ is treated as the original model in which we test that the coefficient vector for x_2 is $0 \in R^2$.

   (iii) Thus "mtest /print;" will display $E \in R^{3\times 3}$ in the true original model.