STT 843: Multivariate Analysis

11. Multivariate Linear Regression (Chapter 7.7)

Guanqun Cao

Department of Statistics and Probability
Michigan State University

Spring 2026

Outline

Unbiased estimate of $\text{var}(\mathbf{y}_{i\cdot}) = \text{var}(\mathbf{e}_{i\cdot}) = \mathbf{\Sigma}$

$$\mathbf{S} = \frac{\mathbf{E}}{n-q-1} = \frac{(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})}{n-q-1} = \frac{\hat{\mathbf{\Xi}}^\top\hat{\mathbf{\Xi}}}{n-q-1}$$
$$= \frac{1}{n-q-1}\left(\mathbf{Y}^\top\mathbf{Y} - \hat{\mathbf{B}}^\top\mathbf{X}^\top\mathbf{Y}\right)$$

Proof:

$$\begin{aligned}
E\{\underset{n\times p}{\hat{\mathbf{\Xi}}}\} &= E\{\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\} \\
&= E\left\{\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)\mathbf{Y}\right\} \\
&= \left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)E\{\mathbf{X}\mathbf{B}+\mathbf{\Xi}\} \\
&= \left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)E\{\mathbf{\Xi}\} \\
&= \mathbf{0}
\end{aligned}$$

$$
\begin{aligned}
E\left\{\hat{\mathbf{e}}_{\cdot i}^{\mathsf{T}}\hat{\mathbf{e}}_{\cdot j}\right\} &= E\left\{\left[\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{y}_{\cdot i}\right]^{\mathsf{T}}\left[\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{y}_{\cdot j}\right]\right\} \\
&= E\left\{\left[\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{e}_{\cdot i}\right]^{\mathsf{T}}\left[\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{e}_{\cdot j}\right]\right\} \\
&= E\{\mathbf{e}_{\cdot i}^{\mathsf{T}}\underbrace{\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)}_{\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}}\mathbf{e}_{\cdot j}\} \\
&= E\left\{\operatorname{tr}\left\{\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{e}_{\cdot j}\mathbf{e}_{\cdot i}^{\mathsf{T}}\right\}\right\} \\
&= \operatorname{tr}\{\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\underbrace{E\left\{\mathbf{e}_{\cdot j}\mathbf{e}_{\cdot i}^{\mathsf{T}}\right\}}_{\sigma_{ij}\mathbf{I}_n}\} \\
&= \sigma_{ij}\operatorname{tr}\left\{\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\right\} \\
&= \sigma_{ij}\left(\operatorname{tr}\{\mathbf{I}_n\} - \operatorname{tr}\left\{\mathbf{X}^{\mathsf{T}}\mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\right\}\right) \\
&= \sigma_{ij}(n - (q+1)) \\
\therefore E\left\{\frac{1}{n-q-1}\hat{\boldsymbol{\Xi}}^{\mathsf{T}}\hat{\boldsymbol{\Xi}}\right\} &= \boldsymbol{\Sigma} = (\sigma_{ij})
\end{aligned}
$$

Note: If $\mathbf{X}$ is not full rank, we can obtain similar results based on $\hat{\mathbf{B}} = \left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\mathbf{Y} \ldots$ we'll leave that discussion for "linear models"

Another note: $\widehat{\mathrm{var}}\{\underset{np\times 1}{\mathbf{e}}\} = \left(\frac{1}{n-q-1}\mathbf{E}\right) \otimes \mathbf{I}_n$

and $E\left\{\left(\frac{1}{n-q-1}\mathbf{E}\right) \otimes \mathbf{I}_n\right\} = \mathbf{\Sigma} \otimes \mathbf{I}_n$

Variance of $\hat{\beta}$ (i.e., $\mathrm{var}\{\mathrm{vec}\,\hat{\mathbf{B}}\}$)

$$
\begin{aligned}
\mathrm{var}\left\{ \begin{matrix} \hat{\beta} \\ rp \times 1 \end{matrix} \right\} &= \mathrm{var}\left\{ \left[(\mathbf{I} \otimes \mathbf{X})^{\mathsf{T}}(\mathbf{I} \otimes \mathbf{X})\right]^{-1} (\mathbf{I} \otimes \mathbf{X})^{\mathsf{T}}\, \mathrm{vec}\,\mathbf{Y} \right\} \\
&= \left[\mathbf{I}_p \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\right] \mathrm{var}\{\mathrm{vec}\,\mathbf{Y}\} \left[\mathbf{I}_p \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\right]^{\mathsf{T}} \\
&= \left[\mathbf{I}_p \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\right] \mathrm{var}\{\mathrm{vec}\,\mathbf{\Xi}\} \left[\mathbf{I}_p \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\right]^{\mathsf{T}} \\
&= \left[\mathbf{I}_p \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\right] (\mathbf{\Sigma} \otimes \mathbf{I}_n) \left[\mathbf{I}_p \otimes \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\right] \\
&= \Sigma \otimes \left[\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\mathbf{I}_n\mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\right] \\
&= \mathbf{\Sigma} \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}
\end{aligned}
$$

Notes:

$$\text{(a) } \widehat{\text{var}}\{\hat{\boldsymbol{\beta}}\} = \overbrace{\left(\frac{1}{n-q-1}\mathbf{E}\right)}^{\text{"S"}} \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}$$

$$\text{(b) } \text{cov}\left\{\hat{\boldsymbol{\beta}}_{\cdot i}, \hat{\boldsymbol{\beta}}_{\cdot j}\right\} = \sigma_{ij}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}$$

(c)

$$
\begin{aligned}
&\operatorname{cov}\left\{\hat{\boldsymbol{\beta}}_{.i}, \hat{\mathbf{e}}_{.j}\right\} \\
&= \operatorname{cov}\left\{\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{y}_{\cdot i}, \left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{y}_{\cdot j}\right\} \\
&= \operatorname{cov}\left\{\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{e}_{\cdot i}, \left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{e}_{\cdot j}\right\} \\
&= \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\sigma_{ij}\mathbf{I}_n\left(\mathbf{I}_n - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right) \\
&= \sigma_{ij}\left[\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}} - \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right] \\
&= \underset{r\times n}{\mathbf{0}}
\end{aligned}
$$

New observation

(d) Estimating mean of $\underset{1\times r}{\mathbf{x}_0^\top}\underset{r\times p}{\mathbf{B}}$

- $\mathbf{x}_0^\top\hat{\mathbf{B}}$ is an unbiased estimator of $\mathbf{x}_0^\top\mathbf{B}$
- $\text{var}\left\{\mathbf{x}_0^\top\hat{\mathbf{B}}\right\} = \mathbf{\Sigma}(\underbrace{\mathbf{x}_0^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{x}_0}_{\text{scalar}})$

(e) Estimating a new observation $\mathbf{y}_0$ using $\mathbf{x}_0$

$\mathbf{y}_0^\top = \mathbf{x}_0^\top\mathbf{B} + \mathbf{e}_0^\top$

- $\mathbf{x}_0^\top \hat{\mathbf{B}}$ is an unbiased estimator of $\mathbf{y}_0$
- $\text{var}\left\{\mathbf{y}_0^\top - \mathbf{x}_0^\top \hat{\mathbf{B}}\right\} \leftarrow$ "forecast error variance"
- Note that

$$\begin{aligned}
\text{cov}\left\{\mathbf{y}_0^\top, \mathbf{x}_0^\top \hat{\mathbf{B}}\right\} &= \text{cov}\left\{\mathbf{e}_0^\top, \mathbf{x}_0^\top \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top (\mathbf{X}\mathbf{B} + \boldsymbol{\Xi})\right\} \\
&= \text{cov}\left\{\mathbf{e}_0^\top, \mathbf{x}_0^\top \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \boldsymbol{\Xi}\right\} \\
&= \underset{p \times p}{\mathbf{0}} \quad \text{since } \mathbf{e}_0^\top \text{ is indep. of } \boldsymbol{\Xi} = \begin{bmatrix} \mathbf{e}_1^\top \\ \vdots \\ \mathbf{e}_{n.}^\top \end{bmatrix}
\end{aligned}$$

So

$$\begin{aligned} \operatorname{var}\left\{\mathbf{y}_0^{\mathsf{T}} - \mathbf{x}_0^{\mathsf{T}}\hat{\mathbf{B}}\right\} &= \operatorname{var}\left\{\mathbf{y}_0^{\mathsf{T}}\right\} + \operatorname{var}\left\{\mathbf{x}_0^{\mathsf{T}}\hat{\mathbf{B}}\right\} - 2\operatorname{cov}\left\{\mathbf{y}_0^{\mathsf{T}}, \mathbf{x}_0^{\mathsf{T}}\hat{\mathbf{B}}\right\} \\ &= \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\cdot\left(\mathbf{x}_0^{\mathsf{T}}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{x}_0\right) + \mathbf{0} \\ &= \boldsymbol{\Sigma}\cdot\left[1 + \mathbf{x}_0^{\mathsf{T}}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{x}_0\right] \end{aligned}$$

MLE's of $\underset{r\times p}{\mathbf{B}}$ and $\boldsymbol{\Sigma}$

Thus far, we have assumed $E\{\mathbf{e}\} = \mathbf{0}$ and $\text{var}\{\mathbf{e}\} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$ vec $\boldsymbol{\Xi}$
If we assume:

$$\underset{np\times 1}{\mathbf{e}} \sim N_{np}\left(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n\right)$$

then MLE's of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are

$$\hat{\mathbf{B}} = \left(\mathbf{X}^\mathsf{T}\mathbf{X}\right)^{-1}\mathbf{X}^\mathsf{T}\mathbf{Y}$$

and

$$\hat{\boldsymbol{\Sigma}}_{p\times p} = \frac{1}{n}\hat{\boldsymbol{\Xi}}^\mathsf{T}\hat{\boldsymbol{\Xi}} = \frac{1}{n}\mathbf{E}$$

where

$$\mathbf{E} \sim W_p(n - q - 1, \boldsymbol{\Sigma})$$

Proof: omitted.

Model Corrected for Means

Rewrite

$$\mathbf{Y} = \underset{n\times r}{\mathbf{X}}\mathbf{B} + \boldsymbol{\Xi}$$

as

$$\underset{n\times p}{\mathbf{Y}_c} = \underset{n\times q}{\mathbf{X}_c}\underset{q\times p}{\mathbf{B}_c} + \boldsymbol{\Xi} \text{ where } q = \#\text{ of predictors } = r-1$$

and

$$\mathbf{Y}_c = \begin{bmatrix} y_{11}-\bar{y}_{\cdot 1} & y_{12}-\bar{y}_{\cdot 2} & \cdots & y_{1p}-\bar{y}_{\cdot p} \\ \vdots & & & \vdots \\ y_{n1}-\bar{y}_{\cdot 1} & y_{n2}-\bar{y}_{\cdot 2} & \cdots & y_{np}-\bar{y}_{\cdot}p \end{bmatrix}$$

$$\mathbf{X}_c = \begin{bmatrix} x_{11}-\bar{x}_{\cdot 1} & \cdots & x_{1q}-\bar{x}_{\cdot q} \\ \vdots & & \vdots \\ x_{n1}-\bar{x}_{\cdot 1} & \cdots & x_{nq}-\bar{x}_{\cdot q} \end{bmatrix}$$

- Then $\hat{\mathbf{B}}_c = \mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$ where $\mathbf{S} = \begin{bmatrix} \mathbf{S}_{yy} & \mathbf{S}_{yx} \\ \mathbf{S}_{xy} & \mathbf{S}_{xx} \end{bmatrix}$ is the sample covariance matrix of the $p+q$ variables $(y_1, \ldots, y_p, x_1, \ldots, x_q)$
- $\hat{\mathbf{Y}} = [\bar{y}_{\cdot 1}\mathbf{1}_n, \ldots, \bar{y}_{\cdot p}\mathbf{1}_n] + \underset{n\times q}{\mathbf{X}_c}\underset{q\times p}{\hat{\mathbf{B}}_c}$

Hypothesis Tests (assuming $\mathbf{e} \sim N_{np}\{\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_r\}$)

$$H_0 : \underset{q\times p}{\mathbf{B}_1} = \mathbf{0} \qquad \text{(Test of overall regression)}$$

$$\text{where } \underset{r\times p}{\mathbf{B}} = \begin{bmatrix} \boldsymbol{\beta}_0^{\mathsf{T}} \\ \mathbf{B}_1 \end{bmatrix} \begin{array}{l} \leftarrow 1 \times p \text{ vector of intercepts} \\ \leftarrow q \times p \end{array}$$

Partition the total SS and CP matrix

$$\begin{aligned}\mathbf{Y}^{\mathsf T}\mathbf{Y} &= (\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})^{\mathsf T}(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}) + \hat{\mathbf{B}}^{\mathsf T}\mathbf{X}^{\mathsf T}\mathbf{Y}^{\mathsf T} \\ &= \mathbf{E} + \hat{\mathbf{B}}^{\mathsf T}\mathbf{X}^{\mathsf T}\mathbf{Y}^{\mathsf T}\end{aligned}$$

To avoid inclusion of $\beta_0^{\mathsf T} = \mathbf{0}^{\mathsf T}$ as part of the null hypothesis, we subtract $n\overline{\mathbf{Y}}\overline{\mathbf{Y}}^{\mathsf T}$:

$$\underbrace{\mathbf{Y}^{\mathsf T}\mathbf{Y} - n\overline{\mathbf{Y}}\overline{\mathbf{Y}}^{\mathsf T}}_{\text{corrected total SS \& CP}} = \underbrace{\mathbf{Y}^{\mathsf T}\mathbf{Y} - \hat{\mathbf{B}}^{\mathsf T}\mathbf{X}^{\mathsf T}\mathbf{Y}}_{\mathbf{E}_{p\times p}} + \underbrace{\hat{\mathbf{B}}^{\mathsf T}\mathbf{X}^{\mathsf T}\mathbf{Y} - n\overline{\mathbf{Y}}\overline{\mathbf{Y}}^{\mathsf T}}_{\mathbf{H}_{p\times p}}$$

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|} = \frac{\left|\mathbf{Y}^{\mathsf T}\mathbf{Y} - \hat{\mathbf{B}}^{\mathsf T}\mathbf{X}^{\mathsf T}\mathbf{Y}\right|}{\left|\mathbf{Y}^{\mathsf T}\mathbf{Y} - n\overline{\mathbf{Y}}\overline{\mathbf{Y}}^{\mathsf T}\right|} \sim \Lambda_{p,\ \underbrace{q,}_{r-1}\ \underbrace{n-q-1}_{n-r}}$$

- $\mathbf{H}$ is "large" when $\hat{\boldsymbol{\beta}}_{ij}$'s are large
- The 4 MANOVA statistics can be calculated as functions of the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}, (\lambda_1, \ldots, \lambda_s)$:
  - Wilks': $\Lambda = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}$
  - Roy's: $\theta = \lambda_1$
  - Pillai's Trace: $V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}$
  - Lawley-Hotelling Trace: $U = \sum_{i=1}^{s} \lambda_i$

Critical values (and p-values) based on approximate F-distributions given on the MANOVA pages

$$s = \min(p, q)$$
$$m = \frac{1}{2}(|q - p| - 1)$$
$$N = \frac{1}{2}(n - q - p - 2)$$

Essential dimensionality of $\underbrace{\mathbf{E}^{-1}\mathbf{H}}_{p\times p}$ is the essential dimensionality of $\mathbf{B}_1$. For example, a single non-zero eigenvalue (i.e., rank of $\mathbf{B}_1$ is 1) $q \times p$ could be due to several causes:

1. $\mathbf{B}_1$ has only one nonzero row $\Rightarrow$ only one of the $\mathbf{X}$'s predicts the $\mathbf{Y}$'s
2. $\mathbf{B}_1$ has only one nonzero column $\Rightarrow$ only one of the $\mathbf{Y}$'s is predicted by the $\mathbf{X}$'s
3. All of the rows of $\mathbf{B}_1$ are linear combinations of each other $\Rightarrow$ $\mathbf{X}$'s act alike in predicting $\mathbf{Y}$'s [or, in other words]
4. All of the columns of $\mathbf{B}_1$ are linear combinations of each other $\Rightarrow$ only one dimension in the $\mathbf{Y}$'s as they relate to $\mathbf{X}$'s

- "Essential dimensionality" of $\mathbf{E}^{-1}\mathbf{H}$ is number of substantially non-zero eigenvalues and takes value less than or equal to $s = \min(p, q)$
- Λ can also be calculated from the partitioned sample covariance matrix of $(y_1, \ldots, y_p, x_1, \ldots, x_q)$

$$\underset{(p+q)\times(p+q)}{\mathbf{S}} = \begin{bmatrix} \underset{p\times p}{\mathbf{S}_{yy}} & \underset{p\times q}{\mathbf{S}_{yx}} \\ \underset{q\times p}{\mathbf{S}_{xy}} & \underset{q\times q}{\mathbf{S}_{xx}} \end{bmatrix}$$

using

$$\frac{|\mathbf{S}|}{|\mathbf{S}_{xx}|\,|\mathbf{S}_{yy}|}$$

which is essentially a test of independence between $\mathbf{y}$ and $\mathbf{x}$ since

Independence of $\mathbf{y}$ and $\mathbf{x} \Rightarrow |\mathbf{S}| = |\mathbf{S}_{yy}|\,|\mathbf{S}_{xx}|$

$$H_0 : \mathbf{B}_a = \mathbf{0} \qquad \text{Tests on a subset of the } \mathbf{x}'s$$

Hypothesis states that the $\mathbf{Y}$'s do not depend on the last h of the $\mathbf{X}$'s. That is,

$$H_0 : \mathbf{B}_{\text{add}} = \mathbf{0}$$

where

$$\underset{r \times p}{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_{\text{red}} \\ \mathbf{B}_{\text{add}} \end{bmatrix} \begin{array}{l} \leftarrow (r-h) \times p \\ \leftarrow h \times p \end{array}$$

Compare SS and CP matrix for full and reduced models

$$\underset{p\times p}{\mathbf{H}_{\text{diff}}} = \hat{\mathbf{B}}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{Y} - \hat{\mathbf{B}}_r^{\mathsf{T}}\mathbf{X}_r^{\mathsf{T}}\mathbf{Y} \quad \leftarrow \text{difference in regression SS and CP}$$

$$\underset{p\times p}{\mathbf{E}_{\text{full}}} = \mathbf{Y}^{\mathsf{T}}\mathbf{Y} - \hat{\mathbf{B}}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{Y} \quad \leftarrow \mathbf{E} \text{ matrix based on full model}$$

Then

$$\begin{aligned}
\Lambda_{x_{q-h+1},\cdots,x_q|x_1,\cdots,x_{q-h}} &= \frac{|\mathbf{E}_{\text{full}}|}{|\mathbf{E}_{\text{full}} + \mathbf{H}_{\mathrm{diff}}|} = \frac{|\mathbf{E}_{\text{full}}|}{|\mathbf{E}_{\mathrm{red}}|} \\
&= \frac{\left|\mathbf{Y}^{\mathsf{T}}\mathbf{Y} - \hat{\mathbf{B}}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\mathbf{Y}\right|}{\left|\mathbf{Y}^{\mathsf{T}}\mathbf{Y} - \hat{\mathbf{B}}_{\mathrm{red}}^{\mathsf{T}}\mathbf{X}_{\mathrm{red}}^{\mathsf{T}}\mathbf{Y}\right|} \\
&\sim \Lambda_{p,h,n-q-1}
\end{aligned}$$

$$\Lambda_{x_{q-h+1},\cdots,x_q|x_1,\cdots,x_{q-h}} = \frac{\left(\frac{\left|\mathbf{Y}^\mathsf{T}\mathbf{Y}-\hat{\mathbf{B}}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{Y}\right|}{\left|\mathbf{Y}^\mathsf{T}\mathbf{Y}-n\overline{\mathbf{Y}}^\mathsf{T}\overline{\mathbf{Y}}\right|}\right)}{\left(\frac{\left|\mathbf{Y}^\mathsf{T}\mathbf{Y}-\hat{\mathbf{B}}_{\mathrm{red}}^\mathsf{T}\mathbf{X}_{\mathrm{red}}^\mathsf{T}\mathbf{Y}\right|}{\left|\mathbf{Y}^\mathsf{T}\mathbf{Y}-n\overline{\mathbf{Y}\mathbf{Y}}^\mathsf{T}\right|}\right)} = \frac{\Lambda_{\mathrm{full}}}{\Lambda_{\mathrm{red}}}$$

$\rightarrow$ makes full vs. reduced testing simple to carry out

- Note: θ, V, and U can be calculated from eigenvalues of $\mathbf{E}_{\mathrm{full}}^{-1}\,\mathbf{H}_{\mathrm{diff}}$ with

$$s = \min(p, h)$$

$$m = \frac{1}{2}(|h-p|-1)$$

$$N = \frac{1}{2}(n-h-p-2)$$

Subset Selection

Finding a subset of the $\mathbf{X}$'s to include in a model

- Forward Selection

(Step 1)
Start with

$$\hat{\mathbf{B}}_i = \begin{bmatrix} \hat{\beta}_{01} & \hat{\beta}_{02} & \cdots & \hat{\beta}_{0p} \\ \hat{\beta}_{i1} & \hat{\beta}_{i2} & \cdots & \hat{\beta}_{ip} \end{bmatrix}$$

and calculate

$$\Lambda_{\mathbf{X}_i} = \frac{\left|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}_i'\mathbf{X}'\mathbf{Y}\right|}{\left|\mathbf{Y}'\mathbf{Y} - n\overline{\mathbf{Y}}\overline{\mathbf{Y}}'\right|} \sim \Lambda_{p,1,n-2}$$

for $i = 1, \ldots, q$. Add the $\mathbf{X}_i$ that minimizes $\Lambda_{\mathbf{X}_i}$ (as long as $\Lambda_{\mathbf{X}_i} < \Lambda_{\alpha,p,1,n-2}$ - stop otherwise)

(Step $j+1, j = 1, 2, \dots$)
Let $\mathbf{X}_1, \dots, \mathbf{X}_j$ be the variables added in previous steps. Calculate

$$\Lambda_{\mathbf{X}_i|\mathbf{X}_1,\dots,\mathbf{X}_j} \sim \Lambda_{p,1,n-j-1}$$

for all $\mathbf{X}_i$ among the $q-j$ remaining candidate variables. For the $\mathbf{X}_i$ that minimizes $\Lambda_{\mathbf{X}_i|\mathbf{X}_1,\dots,\mathbf{X}_j}$:

- add $\mathbf{X}_i$ if $\Lambda_{\mathbf{X}_i|\mathbf{X}_1,\dots,\mathbf{X}_j} < \Lambda_{\alpha,p,1,n-j-1}$
- stop the procedure if $\Lambda_{\mathbf{X}_i|\mathbf{X}_1,\dots,\mathbf{X}_j} > \Lambda_{\alpha,p,1,n-j-1}$

Backward Elimination

Start with all $\mathbf{X}$'s and delete one at a time until the least valuable remaining $\mathbf{X}$ is significant. For the m remaining $\mathbf{X}$'s after a given step, find the $\mathbf{X}_i$ maximizing

$$\Lambda_{\mathbf{x}_i|\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\ldots,\mathbf{x}_m} \sim \Lambda_{p,1,n-m-1}$$

- drop $\mathbf{X}_i$ if $\Lambda_{\mathbf{x}_i|\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\ldots,\mathbf{x}_m} > \Lambda_{\alpha,p,1,n-m-1}$
- stop the procedure if $\mathbf{X}_i$ if $\Lambda_{\mathbf{x}_i|\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\ldots,\mathbf{x}_m} < \Lambda_{\alpha,p,1,n-m-1}$

Stepwise

- Add most significant candidate $\mathbf{X}_i$ if partial Λ is less than critical value
- Then, remove least significant selected $\mathbf{X}_i$ if partial Λ is greater than critical value

Best Subsets

- Choose "best" subset of size ℓ, for $\ell = 1, \ldots, q$, with respect to some criterion (e.g., a multivariate extension of Mallow's C_p, or $\mathrm{tr}\{\mathbf{S}\}$, etc.)