STT 843: Multivariate Analysis

10. Multivariate Linear Regression (Chapter 7.7-1)

Guanqun Cao

Department of Statistics and Probability
Michigan State University

Spring 2026

Outline

Multivariate Multiple Regression

A short review of vec and Kronecker notation

$$\text{Let } \underset{m\times n}{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_{1\cdot}^{\mathsf{T}} \\ \mathbf{a}_{2\cdot}^{\mathsf{T}} \\ \vdots \\ \mathbf{a}_{m\cdot}^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} \begin{array}{c} \mathbf{a}_{\cdot 1} \\ \uparrow \\ \text{an } m\text{-vector} \end{array} & \mathbf{a}_{\cdot 2} & \cdots & \mathbf{a}_{\cdot n} \end{bmatrix}$$

$$\text{vec } \mathbf{A} = \begin{bmatrix} \mathbf{a}_{\cdot 1} \\ \mathbf{a}_{\cdot 2} \\ \vdots \\ \mathbf{a}_{\cdot n} \end{bmatrix}$$

R programming: "$c(A)$" gives vec $\mathbf{A}$

Let $\underset{p\times q}{\mathbf{B}} = (b_{ij})$

- $$\underset{m\times n}{\mathbf{A}} \otimes \underset{p\times q}{\mathbf{B}} = \underbrace{\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}}_{mp\times nq}$$

R: "kronecker (A, B) "gives $\mathbf{A} \otimes \mathbf{B}$

Some properties (without proof)

Assuming that all dimensions are appropriate for matrix multiplication.

$\checkmark$(a)$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$
$\checkmark$(b) $\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^\mathsf{T} \otimes \mathbf{A}\right) \mathrm{vec}\, \mathbf{B}$
(c) $\mathrm{tr}\{\mathbf{AB}\} = \left(\mathrm{vec}\, \mathbf{A}^\mathsf{T}\right)^\mathsf{T} \mathrm{vec}\ \mathbf{B} = (\mathrm{vec}\, \mathbf{A})^\mathsf{T}\ \mathrm{vec}\ \mathbf{B}^\mathsf{T}$
(d) $\mathrm{tr}\{\mathbf{ABCD}\} = \left(\mathrm{vec}\, \mathbf{A}^\mathsf{T}\right)^\mathsf{T} \left(\mathbf{D}^\mathsf{T} \otimes \mathbf{B}\right)\ \mathrm{vec}\ \mathbf{C} = (\mathrm{vec}\, \mathbf{A})^\mathsf{T} \left(\mathbf{B} \otimes \mathbf{D}^\mathsf{T}\right)$ $\mathrm{vec}\ \mathbf{C}^\mathsf{T}$
$\checkmark$(e)$(\mathbf{A} \otimes \mathbf{B})^\mathsf{T} = \mathbf{A}^\mathsf{T} \otimes \mathbf{B}^\mathsf{T}$
$\checkmark$(f)$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$

Univariate Multiple Regression:

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times r}{\mathbf{X}}\,\underset{r\times 1}{\boldsymbol{\beta}}$$

- Assume $E\{\mathbf{e}\} = \mathbf{0}$ and $\text{var}\{\mathbf{e}\} = \sigma^2\mathbf{I}_n$. Then

$$\hat{\boldsymbol{\beta}}$$
$$\uparrow$$
O.L.S.
estimator

is B.L.U.E. for $\boldsymbol{\beta}$.

- Note: we'll use q to denote the # of $\mathbf{X}$'s and $r = q + 1$ to denote the # of columns in the $\mathbf{X}$ matrix when using an intercept

Multivariate Multiple Regression

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\Xi}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_{1\cdot}^{\mathsf{T}} \\ \vdots \\ \mathbf{y}_{n\cdot}^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{\cdot 1} & \cdots & \mathbf{y}_{\cdot p} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_{\cdot 1} & \boldsymbol{\beta}_{\cdot 2} & \cdots & \boldsymbol{\beta}_{\cdot p} \end{bmatrix}$$

$$\boldsymbol{\Xi} = \begin{bmatrix} \mathbf{e}_{1\cdot}^{\mathsf{T}} \\ \vdots \\ \mathbf{e}_{n\cdot}^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{\cdot 1} & \cdots & \mathbf{e}_{\cdot p} \end{bmatrix}$$

- Note that

$$\underset{n\times 1}{\mathbf{y}_{\cdot j}} = \underset{n\times r}{\mathbf{X}}\underset{r\times 1}{\boldsymbol{\beta}_{\cdot j}} + \underset{n\times 1}{\mathbf{e}_{\cdot \mathbf{j}}}$$

- Assume $E\{\boldsymbol{\Xi}\} = \mathbf{0}$, $\text{var}\{\mathbf{e}_i\} = \underset{p\times p}{\boldsymbol{\Sigma}}$, and $\text{cov}\{\mathbf{e}_i, \mathbf{e}_{k\cdot}\} = \underset{p\times p}{\mathbf{0}}$ for all $i \neq k$
- Question: Is $\underset{r\times p}{\hat{\mathbf{B}}} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}$ a B.L.U.E.?

Rewrite model:

$$\text{vec}\,\mathbf{Y} = \text{vec}(\mathbf{XB}) + \text{vec}(\boldsymbol{\Xi})$$

$$= \underbrace{(\mathbf{I}_p \otimes \mathbf{X})}_{\uparrow} \quad \underbrace{\text{vec}\ \mathbf{B}}_{\equiv \underset{pr\times 1}{\boldsymbol{\beta}}} + \underbrace{\text{vec}\ (\boldsymbol{\Xi})}_{\equiv \underset{np\times 1}{\mathbf{e}}}$$

$$\text{rank} \;= pr$$
$$\text{when}\;\; \text{rank}(\mathbf{X}) = r$$

Note: $E\{\mathbf{e}\} = \mathbf{0}$

$$
\begin{aligned}
&\text{var}\{\mathbf{e}\} \\
&= \text{var}\left\{\begin{pmatrix} \mathbf{e}_{\cdot 1} \\ \vdots \\ \mathbf{e}_{\cdot p} \end{pmatrix}\right\} = \begin{bmatrix} \sigma_{11}\mathbf{I}_n & \sigma_{12}\mathbf{I}_n & \cdots & \sigma_{1p}\mathbf{I}_n \\ \sigma_{21}\mathbf{I}_n & \sigma_{22}\mathbf{I}_n & \cdots & \sigma_{2p}\mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}\mathbf{I}_n & \sigma_{p2}\mathbf{I}_n & \cdots & \sigma_{pp}\mathbf{I}_n \end{bmatrix} = \underset{p\times p}{\boldsymbol{\Sigma}} \otimes \mathbf{I}_n
\end{aligned}
$$

- Since $\text{var}\{\underset{np\times 1}{\mathbf{e}}\}$ does not take the form $\sigma^2\mathbf{I}_{np}$, the B.L.U.E. for β will be the G.L.S. estimator for β (which depends on the unknown $\boldsymbol{\Sigma}$) BUT...

$$\hat{\beta} = \left[\underbrace{\left(\mathbf{I}_p \otimes \underset{n\times p}{\mathbf{X}}\right)^{\mathsf{T}}}_{\left(\mathbf{I}_p\otimes\mathbf{X}^{\mathsf{T}}\right)} \underbrace{\left(\boldsymbol{\Sigma}\otimes\mathbf{I}_n\right)^{-1}}_{\left(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_n\right)} \left(\mathbf{I}_p \otimes \underset{n\times p}{\mathbf{X}}\right)\right]^{-1}$$

$$\times \left(\mathbf{I}_p\otimes\mathbf{X}\right)^{\mathsf{T}}\left(\boldsymbol{\Sigma}\otimes\mathbf{I}_n\right)^{-1}\operatorname{vec}\mathbf{Y}$$

$$= \left[\left(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{X}^{\mathsf{T}}\right)\left(\mathbf{I}_p\otimes\mathbf{X}\right)\right]^{-1}\left(\mathbf{I}_p\otimes\mathbf{X}^{\mathsf{T}}\right)\left(\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_n\right)\operatorname{vec}\mathbf{Y}$$

$$
\begin{aligned}
&= \left[\boldsymbol{\Sigma}^{-1} \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)\right]^{-1} \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}^{\mathsf{T}}\right) \text{ vec } \mathbf{Y} \quad [\text{ by prop (a) }] \\
&= \left(\boldsymbol{\Sigma} \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\right) \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}^{\mathsf{T}}\right) \text{ vec } \mathbf{Y} \quad [\text{ by prop } (\mathrm{f})] \\
&= \left(\mathbf{I}_p \otimes \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\right) \text{ vec } \mathbf{Y} \quad [\text{ by prop (a) }] \\
&\Rightarrow \hat{\mathbf{B}} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1} \mathbf{X}^{\mathsf{T}}\mathbf{Y} \quad [\text{ by prop (b) }]
\end{aligned}
$$

- O.L.S. = G.L.S. is BLUE! (Even when $\boldsymbol{\Sigma}$ is unknown)
- Despite the fact that the p variables $y_{i1}, \ldots, y_{ip}$ are correlated, all the info needed to estimate $\underset{r \times 1}{\boldsymbol{\beta}_{.i}}$ is found in $\mathbf{y}_{.i}$ only. That is, multivariate regression coefficient matrix $\underset{r \times p}{\hat{\mathbf{B}}}$ can be formed by pasting together the p columns from p separate univariate regressions (as long as each regression uses the same predictors $\underset{n \times r}{\mathbf{X}}$)
- But all $\hat{\beta}_{ij}$ in B are intercorrelated ... must take multivariate approach to inference

Assumptions for Multivariate Multiple Regression

Model is: $\underset{n\times p}{\mathbf{Y}} = \underset{n\times r}{\mathbf{X}}\underset{r\times p}{\mathbf{B}}$ or $\operatorname{vec}\mathbf{Y} = (\mathbf{I}\otimes\mathbf{X})\underbrace{\operatorname{vec}\mathbf{B}}_{=''\boldsymbol{\beta}''} + \operatorname{vec}\boldsymbol{\Xi}$

Assumptions:

1. $E\{\mathbf{Y}\} = \mathbf{X}\mathbf{B}$ or $E\{\boldsymbol{\Xi}\} = \mathbf{0}$
2. $\operatorname{var}\{\operatorname{vec}\mathbf{Y}\} = \operatorname{var}\{\operatorname{vec}\boldsymbol{\Xi}\} = \boldsymbol{\Sigma}\otimes\mathbf{I}_n$

(That is, $\operatorname{var}\{\mathbf{y}_i\} = \boldsymbol{\Sigma}$ for all $i = 1,\dots,n$ and $\operatorname{cov}\{\mathbf{y}_i,\mathbf{y}_j\} = \underset{p\times p}{\mathbf{0}}$ for all $i \neq j$)

Some properties of $\hat{\mathbf{B}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$

1. $\hat{\mathbf{B}}$ is called the "least squares estimator" because it "minimizes" $\underset{p\times p}{\mathbf{E}} = \hat{\boldsymbol{\Xi}}^\top\hat{\boldsymbol{\Xi}} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$ (where $\mathbf{E}$ is an "error matrix" analogous to the $\mathbf{E}$ matrix in MANOVA). Matrix is "minimized" in several senses:

(a) Let $\tilde{\mathbf{B}}$ be some other estimate of $\mathbf{B}$.
Then,

$$(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}})^T(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}}) = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^T(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + \mathbf{A}$$

where $\mathbf{A}$ is a positive definite matrix
(b) $\mathbf{B} = \hat{\mathbf{B}}$ minimizes $\mathrm{tr}\left\{(\mathbf{Y} - \mathbf{X}\mathbf{B})^T(\mathbf{Y} - \mathbf{X}\mathbf{B})\right.$
(c) $\mathbf{B} = \hat{\mathbf{B}}$ minimizes $\left|(\mathbf{Y} - \mathbf{X}\mathbf{B})^T(\mathbf{Y} - \mathbf{X}\mathbf{B})\right|$

2. Let $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} = \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{Y}$ be predicted values and $\hat{\boldsymbol{\Xi}} = \mathbf{Y} - \hat{\mathbf{Y}} = \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{Y}$ be residuals Then

(a) Residuals are perpendicular to the columns of $\mathbf{X}$

$$\rightarrow \mathbf{X}^{\mathsf{T}}\hat{\boldsymbol{\Xi}} = \mathbf{X}^{\mathsf{T}}\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{Y} = \underset{r\times p}{\mathbf{0}}$$

(b) Residuals are perpendicular to the columns of $\hat{\mathbf{Y}}$

$$\rightarrow \hat{\mathbf{Y}}^{\mathsf{T}}\hat{\boldsymbol{\Xi}} = \hat{\mathbf{B}}^{\mathsf{T}}\mathbf{X}^{\mathsf{T}}\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\right)\mathbf{Y} = \underset{p\times p}{\mathbf{0}}$$

(c) Total sum of squares and cross products ("Total SS and CP") can be partitioned as:

$$\mathbf{Y}^{\mathsf{T}}\mathbf{Y} = (\hat{\mathbf{Y}} + \hat{\boldsymbol{\Xi}})^{\mathsf{T}}(\hat{\mathbf{Y}} + \hat{\boldsymbol{\Xi}})$$

$$\underbrace{\mathbf{Y}^{\mathsf{T}}\mathbf{Y}}_{\substack{\uparrow \\ \text{total}}} = \underbrace{\hat{\mathbf{Y}}^{\mathsf{T}}\hat{\mathbf{Y}}}_{\substack{\uparrow \\ \hat{Y}^{\mathsf{T}}\hat{\mathbf{Y}} \\ \text{matrix}}} + \underbrace{\hat{\boldsymbol{\Xi}}^{\mathsf{T}}\hat{\boldsymbol{\Xi}}}_{\substack{\text{SS\&CP} \\ \text{matrix}}}$$

3. $\hat{\mathbf{B}}$ is B.L.U.E. for $\mathbf{B}$
   - Minimum variance estimator among all unbiased estimators
   - If columns of $\boldsymbol{\Xi}$ are normal, $\hat{\mathbf{B}}$ is B.U.E.
4. Elements of $\hat{\mathbf{B}}$ are intercorrelated

$$\underset{r\times p}{\mathbf{B}} = \begin{bmatrix} \hat{\beta}_{01} & \hat{\beta}_{02} & \cdots & \hat{\beta}_{0p} \\ \hat{\beta}_{11} & \hat{\beta}_{12} & \cdots & \hat{\beta}_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{q1} & \hat{\beta}_{q2} & \cdots & \hat{\beta}_{qp} \end{bmatrix}$$

where $r = q + 1$

- $\hat{\beta}$'s in each row are correlated due to correlation in y
- $\hat{\beta}$'s in each column are correlated due to correlation in $\mathbf{x}$