Topic 2: Linear models with multivariate response

Univariate linear models and ANOVA

Recall the univariate linear model: for observations $i = 1, 2, \ldots, n$, let the response variable $Y_i = \boldsymbol{x}_i \boldsymbol{\beta} + \epsilon_i$, for predictor row vector $\boldsymbol{x}_i^\top \in \mathbb{R}^k$ assumed fixed and known, coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$ fixed and unknown, and $\epsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$. In matrix form, $\boldsymbol{Y} = \begin{pmatrix} Y_1 & Y_2 & \cdots & Y_n \end{pmatrix}^\top$ and $X = \begin{pmatrix} x_1^\top & x_2^\top & \cdots & x_n^\top \end{pmatrix}^\top \in \mathcal{M}_{n,k}$. We will assume that X contains an intercept. Then,

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, I_n \sigma^2)$. The MLE for $\boldsymbol{\beta}$ requires us to minimise

$$\sum_{i=1}^{n} (Y_i - \boldsymbol{x}_i \boldsymbol{\beta})^2 = |\boldsymbol{Y} - X\boldsymbol{\beta}|^2 = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

and, after some vector calculus, we get

$$\hat{\boldsymbol{\beta}} = \left(X^\top X\right)^{-1} X^\top \boldsymbol{Y}$$

with

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = \left(X^\top X\right)^{-1} X^\top \operatorname{Var}(\boldsymbol{Y}) X \left(X^\top X\right)^{-1} = \left(X^\top X\right)^{-1} \sigma^2$$

Furthermore, we can consider projection matrices $A = I_n - X\left(X^\top X\right)^{-1} X^\top$ and $B = X\left(X^\top X\right)^{-1} X^\top - \mathbf{1}_n \left(\mathbf{1}_n^\top \mathbf{1}_n\right)^{-1} \mathbf{1}_n^\top$ with

$$A\boldsymbol{Y} = \boldsymbol{Y} - X\left\{\left(X^\top X\right)^{-1} X^\top \boldsymbol{Y}\right\} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$$

the residual vector and

$$B\boldsymbol{Y} = X\left\{(X^\top X)^{-1} X^\top\right\} \boldsymbol{Y} - \mathbf{1}_n (\mathbf{1}_n^\top \mathbf{1}_n)^{-1} \mathbf{1}_n^\top \boldsymbol{Y} = \hat{\boldsymbol{Y}} - \mathbf{1}_n \bar{\boldsymbol{Y}}$$

the vector of fitted values over and above the mean, and observe that

$$\begin{aligned}
\operatorname{Cov}(A\boldsymbol{Y}, B\boldsymbol{Y}) =& A \operatorname{Var}(\boldsymbol{Y}) B^\top = \sigma^2 A B^\top \\
=& X\left(X^\top X\right)^{-1} X^\top - X\left(X^\top X\right)^{-1} X^\top X \left(X^\top X\right)^{-1} X^\top \\
& - \mathbf{1}_n \left(\mathbf{1}_n^\top \mathbf{1}_n\right)^{-1} \mathbf{1}_n^\top + X\left(X^\top X\right)^{-1} X^\top \mathbf{1}_n \left(\mathbf{1}_n^\top \mathbf{1}_n\right)^{-1} \mathbf{1}_n^\top \\
=& \frac{1}{n}\left(X\left(X^\top X\right)^{-1} X^\top \mathbf{1}_n - \mathbf{1}_n\right) \mathbf{1}_n^\top = \mathbf{0}
\end{aligned}$$

if X contains an intercept effect. Then, $\mathrm{SSE} = \boldsymbol{Y}^\top A \boldsymbol{Y} \sim \sigma^2 \chi^2_{n-k}$ and $\mathrm{SSA} = \boldsymbol{Y}^\top B \boldsymbol{Y} \sim \sigma^2 \chi^2_{k-1}$, independent, letting us set up $F = \frac{\mathrm{SSA}\,/(k-1)}{\mathrm{SSE}\,/(n-k)} \sim F_{k-1,n-k}$, etc.

Optional reading

For an alternate presentation of these concepts, read the following:

Härdle W.K., Simar L. (2012) Regression Models. In: Applied Multivariate Statistical Analysis. Springer, Berlin, Heidelberg

- Section 8.1.2

All readings are available from the course Leganto reading list. Please keep in mind that you will need to be logged into Moodle to access the Leganto reading list.

Multivariate Linear Model and MANOVA

How do we generalise the previous slide's ideas to a multivariate response? That is, suppose that we observe the following response matrix:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1^\top \\ \boldsymbol{Y}_2^\top \\ \vdots \\ \boldsymbol{Y}_n^\top \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1p} \\ Y_{21} & Y_{22} & \cdots & Y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{np} \end{pmatrix} \in \mathcal{M}_{n,p}$$

with $\boldsymbol{x}_i$ and X as before, and

$$\boldsymbol{Y}_i^\top = \boldsymbol{x}_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i^\top$$

where $\boldsymbol{\beta} \in \mathcal{M}_{k,p}$, and $\boldsymbol{\epsilon}_i \sim N_p(\boldsymbol{0}, \Sigma), \Sigma \in \mathcal{M}_{p,p}$ symmetric positive definite. In matrix form,

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{E}$$

where

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{\epsilon}_1 & \boldsymbol{\epsilon}_2 & \cdots & \boldsymbol{\epsilon}_n \end{pmatrix}^\top \in \mathcal{M}_{n,p}$$

To express the properties of these estimators concisely, we require *Kronecker products* that were previously introduced in optional materials. Kronecker product of two matrices $A \in \mathcal{M}_{m,n}$ *and* $B \in \mathcal{M}_{p,q}$ is denoted by $A \otimes B$ and is defined (in block matrix notation) as

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}. \tag{1.13}$$

A related useful concept is the $\vec{\square}$ operation on a matrix $A \in \mathcal{M}_{m,n}$. This operation creates a vector $\vec{A} \in \mathbb{R}^{mn}$ which is composed by stacking the n columns of the matrix $A \in \mathcal{M}_{m,n}$ under each other (the second below the first etc).

From the derivations given below, we get:

$$\hat{\boldsymbol{\beta}} = \left(X^\top X\right)^{-1} X^\top \boldsymbol{Y}$$

and

$$\operatorname{Var}\left(\overrightarrow{\hat{\boldsymbol{\beta}}^\top}\right) = \left(X^\top X\right)^{-1} \otimes \Sigma$$

or

$$\operatorname{Var}(\vec{\hat{\boldsymbol{\beta}}}) = \Sigma \otimes \left(X^\top X\right)^{-1}$$

Projection matrices A and B still work, and we can write $\mathrm{SSE} = \boldsymbol{Y}^\top A \boldsymbol{Y} \sim W_p(\Sigma, p(n-k-1))$ and $\mathrm{SSA} = \boldsymbol{Y}^\top B \boldsymbol{Y} \sim W_p(\Sigma, p(k-1))$. Notice that they are now matrices.

> The following derivation details are optional and will not be assessed.

The following basic properties of Kronecker products will be used:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$

$$(A + B) \otimes C = A \otimes C + B \otimes C$$

$$(A \otimes B)^\top = A^\top \otimes B^\top$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

(whenever the corresponding matrix products and inverses exist)

$$\operatorname{tr}(A \otimes B) = \operatorname{tr}(A) \operatorname{tr}(B)$$

$$|A \otimes B| = |A|^p |B|^m$$

(in case $A \in \mathcal{M}_{m,m}, B \in \mathcal{M}_{p,p}$).

For matrices A, B and C (of suitable dimensions) it holds:

$$\overrightarrow{ABC} = (C^\top \otimes A)\vec{B}$$

Then, we can write $\vec{\boldsymbol{E}} \sim N_{np}(\mathbf{0}, \Sigma \otimes I_n)$ or $\overrightarrow{\boldsymbol{E}^\top} \sim N_{np}(\mathbf{0}, I_n \otimes \Sigma)$, and

$$\vec{\boldsymbol{Y}} \sim N_{np}\left(\left\{\boldsymbol{\beta}^\top \otimes I_n\right\} \vec{X}, \Sigma \otimes I_n\right)$$

or

$$\overrightarrow{\boldsymbol{Y}^\top} \sim N_{np}\left(\left\{I_n \otimes \boldsymbol{\beta}^\top\right\} \overrightarrow{X^\top}, I_n \otimes \Sigma\right)$$

MLE is equivalent to the OLS problem minimising $\sum_{i=1}^n \operatorname{tr}\left\{(\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta})(\boldsymbol{Y}_i - \boldsymbol{x}_i\boldsymbol{\beta})^\top\right\} =$

$\mathrm{tr}\left\{(\boldsymbol{Y} - X\boldsymbol{\beta})^\top(\boldsymbol{Y} - X\boldsymbol{\beta})\right\}.$

Then,

$$\hat{\boldsymbol{\beta}} = \left(X^\top X\right)^{-1} X^\top \boldsymbol{Y}$$

again, with

$$\begin{aligned}
\mathrm{Var}\left(\overrightarrow{\hat{\boldsymbol{\beta}}^\top}\right) &= \mathrm{Var}(\overrightarrow{Y^\top X \left(X^\top X\right)^{-1}}) = \mathrm{Var}\left\{\left(\left(X^\top X\right)^{-1} X^\top \otimes I_p\right) \overrightarrow{Y^\top}\right\} \\
&= \left(\left(X^\top X\right)^{-1} X^\top \otimes I_p\right) (I_p \otimes \Sigma) \left(\left(X^\top X\right)^{-1} X^\top \otimes I_p\right)^\top \\
&= \left(\left(X^\top X\right)^{-1} X^\top \otimes I_p\right) \left(\left(X^\top X\right)^{-1} X^\top \otimes \Sigma\right)^\top \\
&= \left(X^\top X\right)^{-1} \otimes \Sigma
\end{aligned}$$