SMIII lectures

Week 1

Linear Regression

Multiple Linear Regression

Notation

- We will use the convention of representing random variables by uppercase letters, e.g. Y, and realisations of random variables by the corresponding lowercase letters, e.g. y.
- In this course we will make extensive use of random vectors and occasional use of random matrices.
- Throughout, we will consider variables for which means and variances exist.

Random vectors

Definition 1.1

A random vector is a vector of random variables. For example,

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}.$$

For the random vector $\boldsymbol{Y}$ we define the mean vector, $\boldsymbol{\eta}$, by

$$\boldsymbol{\eta} = E(\boldsymbol{Y}) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix}.$$

The variance matrix is defined by

$$\mathrm{Var}(\boldsymbol{Y}) = \Sigma = [\sigma_{ij}]$$

where

$$\sigma_{ij} = \begin{cases} \mathrm{cov}(Y_i, Y_j) & \text{for } i \neq j, \\ \mathrm{var}(Y_i) & \text{for } i = j. \end{cases}$$

$\square$

Random matrix

A random matrix can also be defined to be a matrix of random variables,

$$\boldsymbol{\mathcal{Y}} = [Y_{ij}]$$

and we will use the convention

$$E(\boldsymbol{\mathcal{Y}}) = [E(Y_{ij})].$$

Note that we will not need to define the variance structure for random matrices.

Linear transformations

Lemma 1.1

Suppose $\boldsymbol{Y}$ is a random vector with $E(\boldsymbol{Y}) = \boldsymbol{\eta}$ and $\mathrm{Var}(\boldsymbol{Y}) = \Sigma$ and let $A_{m\times n}$ and $\boldsymbol{b}_{m\times 1}$ be fixed. Then

$$E(A\boldsymbol{Y} + \boldsymbol{b}) = A\boldsymbol{\eta} + \boldsymbol{b} \text{ and } \mathrm{Var}(A\boldsymbol{Y} + \boldsymbol{b}) = A\Sigma A^T.$$

If $\boldsymbol{\mathcal{Y}}$ is a random matrix and A is a fixed matrix then

$$E(A\boldsymbol{\mathcal{Y}}) = AE(\boldsymbol{\mathcal{Y}}).$$

$\square$

Normal distribution

In this course, we will use the notation,

$$\boldsymbol{Y} \sim N_r(\boldsymbol{\mu}, \Sigma)$$

to indicate that the r-dimensional random vector $\boldsymbol{Y}$ has the r-dimensional **multivariate normal distribution** with mean vector $\boldsymbol{\mu}$ and variance matrix Σ.

Normal distribution results

Lemma 1.2

If $\boldsymbol{Y} \sim N_r(\boldsymbol{\mu}, \Sigma)$ and $A_{k\times r}$ and $\boldsymbol{b}_{k\times 1}$ are fixed then

$$A\boldsymbol{Y} + \boldsymbol{b} \sim N_k(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^T).$$

If $\boldsymbol{Y} \sim N_r(\boldsymbol{\mu}, \Sigma)$ and $\boldsymbol{a}_{r\times 1}$ is fixed, then

$$\boldsymbol{a}^T\boldsymbol{Y} \sim N(\boldsymbol{a}^T\boldsymbol{\mu}, \boldsymbol{a}^T\Sigma\boldsymbol{a}).$$

$\square$

Multiple regression

- The regression model is used to model the dependence between a predictor variable x and a response variable Y.
- In general, there may be several predictor variables $x_1, x_2, \ldots, x_r$ and single response Y.
- In this case the *multiple regression* model may be used to model the simultaneous influence of the predictors.

Notation

Consider data,

$$
\begin{aligned}
&(y_1, x_{11}, x_{12}, \ldots, x_{1r}) \\
&(y_2, x_{21}, x_{22}, \ldots, x_{2r}) \\
&\ldots \\
&(y_n, x_{n1}, x_{n2}, \ldots, x_{nr})
\end{aligned}
$$

Multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_r x_{ir} + e_i$$

where $e_1, e_2, \ldots, e_n$ are realisations of independent random variables $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ with

$$E(\mathcal{E}_i) = 0 \text{ and } \operatorname{var}(\mathcal{E}_i) = \sigma^2.$$

Alternative forms

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_r x_{ir} + e_i$$

with $e_1, e_2, \ldots, e_n$ *i.i.d.* $N(0, \sigma^2)$ as an abbreviation for the random variable formulation given above.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_r x_{ir} + \mathcal{E}_i$$

Matrix Formulation

Let

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \; X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1r} \\ 1 & x_{21} & x_{22} & \dots & x_{2r} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nr} \end{pmatrix}, \; \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{pmatrix}, \; \boldsymbol{\mathcal{E}} = \begin{pmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \vdots \\ \mathcal{E}_n \end{pmatrix}.$$

The multiple regression model can then be formulated as

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{e}$$

or, in terms of random variables,

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

with

$$E(\boldsymbol{\mathcal{E}}) = \mathbf{0} \text{ and } \operatorname{Var}(\boldsymbol{\mathcal{E}}) = \sigma^2 I_{n\times n}.$$

The additional assumption of normality is then formulated as

$$\boldsymbol{\mathcal{E}} \sim N_n(\mathbf{0}, \sigma^2 I).$$

Linear Independence

Definition 1.2

A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\}$ is said to be **linearly independent** if

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \ldots + \alpha_p \boldsymbol{v}_p = \mathbf{0} \quad \Rightarrow \quad \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0.$$

Otherwise it is said to be **linearly dependent**.

$\square$

Remark

When $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ are linearly dependent it means that one of the $\boldsymbol{v}_i$'s is expressible as a linear combination of the remaining $\boldsymbol{v}$'s.

Identifiability

Consider the multiple regression model

$$\boldsymbol{y} = X\beta + \boldsymbol{e}.$$

We require that the columns of X be linearly independent.

Linear least squares

Definition 1.3

The least squares estimate, $\hat{\beta}$ is the vector that minimises the sum of squares

$$Q(\beta) = \|\boldsymbol{y} - X\beta\|^2.$$

The variance σ^2 is estimated by

$$\begin{aligned} s_e^2 &= \frac{1}{n-p} \sum_{i=1}^{n} \{y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_r x_{ir})\}^2 \\ &= \frac{1}{n-p} \|\boldsymbol{y} - X\hat{\beta}\|^2. \end{aligned}$$

where $p = r + 1$ is the number of columns of X.

$\square$

Theorem 1.1

If the columns of X are linearly independent then the least squares estimates are given uniquely by

$$\hat{\beta} = (X^T X)^{-1} X^T \boldsymbol{y}.$$

$\square$

Fitted values

The vector of fitted values is defined by

$$\hat{\boldsymbol{\eta}} = X\hat{\beta} = X(X^T X)^{-1} X^T \boldsymbol{y} = P\boldsymbol{y}$$

where $P = X(X^T X)^{-1} X^T$.

Alternative notation

- Alternative notation: $H = X(X^T X)^{-1} X^T$.
- For reasons to be discussed later, the $n \times n$ matrix P is called an orthogonal projection matrix.
- The elementary statistical properties of $\hat{\beta}$ and s_e^2 are summarised in the following theorem.

Theorem 1.2

Suppose

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

where

$$E(\boldsymbol{\mathcal{E}}) = \mathbf{0} \text{ and } \mathrm{Var}(\boldsymbol{\mathcal{E}}) = \sigma^2 I.$$

- $E(\hat{\beta}) = \beta$.
- $\mathrm{Var}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$
- $E(s_e^2) = \sigma^2$.

$\square$

Theorem 1.3

Suppose

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

where

$$E(\boldsymbol{\mathcal{E}}) = \mathbf{0} \text{ and } \mathrm{Var}(\boldsymbol{\mathcal{E}}) = \sigma^2 I.$$

If $\boldsymbol{\mathcal{E}} \sim N_n(\mathbf{0}, \sigma^2 I)$, then:

- $\hat{\beta} \sim N_p(\beta, \sigma^2(X^T X)^{-1})$.
- $\dfrac{(n-p)s_e^2}{\sigma^2} \sim \chi^2_{n-p}$ independently of $\hat{\beta}$.

$\square$