5100 Methods Notes

Bookmark:

Key LM Results

A General Linear Model (GLM)

Suppose

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

- $\mathbf{y} \in \mathbb{R}^n$ is the response vector,
- $\mathbf{X}$ is an $n \times p$ matrix of known (fixed) constants,
- $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown parameter vector, and
- $\boldsymbol{\varepsilon}$ is a vector of unobserved random errors satisfying

$$\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \qquad \mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}.$$

The model is called a *linear model* because the mean of the response vector is linear in the unknown parameter vector:

$$\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}.$$

Ordinary Least Squares (OLS) Estimation

Assume

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \qquad \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}.$$

Then

$$\mathbb{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \in \mathcal{C}(\mathbf{X}),$$

where $\mathcal{C}(\mathbf{X})$ denotes the column space of $\mathbf{X}$.

To estimate $\mathbb{E}(\mathbf{y})$, we consider vectors of the form $\mathbf{X}\hat{\boldsymbol{\beta}}$.

Thus, estimating $\mathbb{E}(\mathbf{y})$ amounts to finding the vector in $\mathcal{C}(\mathbf{X})$ that is closest to $\mathbf{y}$.

Let $\mathcal{N}(\mathbf{X}^\top)$ denote the null space of $\mathbf{X}^\top$.

Then $\mathcal{C}(\mathbf{X})$ and $\mathcal{N}(\mathbf{X}^\top)$ are orthogonal complements:

$$\mathcal{N}(\mathbf{X}^\top) \perp \mathcal{C}(\mathbf{X}).$$

The null space of a matrix $\mathbf{A}$ is defined as

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}\}.$$

Least Squares Estimate (LSE)

An estimate $\hat{\boldsymbol{\beta}}$ is a *least squares estimate* (LSE) of $\boldsymbol{\beta}$ if $\mathbf{X}\hat{\boldsymbol{\beta}}$ is the vector in $\mathcal{C}(\mathbf{X})$ that is closest to $\mathbf{y}$.

Equivalently,

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Define the error sum of squares:

$$Q(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Identifying the LSE

There are two equivalent approaches:

- **Algebraic**: solving the normal equations
- **Geometric**: orthogonal projection of $\mathbf{y}$ onto $\mathcal{C}(\mathbf{X})$

Normal Equations

Expand the objective function:

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

Taking derivatives and setting the gradient equal to zero yields

$$\nabla Q(\boldsymbol{\beta}) = -2\mathbf{X}^\top\mathbf{y} + 2\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = \mathbf{0}.$$

This leads to the **normal equations**:

$$\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^\top\mathbf{y}.$$

Solutions to the Normal Equations

If $\operatorname{rank}(\mathbf{X}) = p$, then $\mathbf{X}^\top\mathbf{X}$ is invertible and the unique solution is

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

If $\operatorname{rank}(\mathbf{X}) < p$, the normal equations have infinitely many solutions.

In this case, $\hat{\boldsymbol{\beta}}$ may not be unique, but

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

is unique.

Geometric Approach

Let $\mathbf{P_X}$ denote the orthogonal projection matrix onto $\mathcal{C}(\mathbf{X})$:

$$\mathbf{P_X} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top,$$

where $(\mathbf{X}^\top\mathbf{X})^-$ is any generalized inverse.

Properties

- $\mathbf{P_X}$ is idempotent:

$$\mathbf{P}_\mathbf{X}^2 = \mathbf{P_X}.$$

- $\mathbf{P_X}$ projects onto $\mathcal{C}(\mathbf{X})$.
- $\mathbf{P_X}$ is symmetric:

$$\mathbf{P}_\mathbf{X}^\top = \mathbf{P_X}.$$

- $\mathbf{P_X}\mathbf{X} = \mathbf{X}$ and $\mathbf{X}^\top\mathbf{P_X} = \mathbf{X}^\top$.
- $\operatorname{rank}(\mathbf{X}) = \operatorname{rank}(\mathbf{P_X}) = \operatorname{tr}(\mathbf{P_X})$.

Fitted Values and Residuals

An estimate $\hat{\boldsymbol{\beta}}$ is a least squares estimate if and only if

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P_X}\mathbf{y}.$$

The OLS estimator of $\mathbb{E}(\mathbf{y})$ is

$$\hat{\mathbf{y}} = \mathbf{P_X}\mathbf{y}.$$

The residual vector is

$$\hat{\boldsymbol{e}} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{P_X})\mathbf{y}.$$

Note that

$$\hat{\boldsymbol{e}} \in \mathcal{N}(\mathbf{X}^\top).$$

Since $\mathcal{C}(\mathbf{X})$ and $\mathcal{N}(\mathbf{X}^\top)$ are orthogonal complements, we obtain the unique decomposition

$$\mathbf{y} = \hat{\mathbf{y}} + \hat{\boldsymbol{e}}.$$

ANOVA Decomposition for the Linear Model

Suppose y is $n \times 1$, X is $n \times p$ with rank $r \leq p$, β is $p \times 1$, and ε is $n \times 1$.
We assume the model given in (1):

$$y = X\beta + \varepsilon.$$

Then, the ANOVA table is:

Source	df	Sum of Squares
Model	r	$\hat{y}^\top \hat{y} = y^\top P_X y$
Residual	$n-r$	$\hat{e}^\top \hat{e} = y^\top (I - P_X) y$
Total	$n-1$	$y^\top y = y^\top I y$

Starting on estimability

For any $q \times n$ matrix A, $AE(y)$ is a linear function of $E(y)$.

For any $q \times n$ matrix A, the OLS estimator of

$$AE(y) = AX\beta$$

is

$$A[\text{OLS Estimator of } E(y)] = A\hat{y} = AP_X y = AX(X^\top X)^- X^\top y.$$

Note that

$$AE(y) = AX\beta$$

is automatically a linear function of β of the form

$$C\beta,$$

where

$$C = AX.$$

If C is any $q \times p$ matrix, we say that the linear function of β given by $C\beta$ is **estimable** if and only if

$$C = AX$$

for some $q \times n$ matrix A.

The OLS estimator of an estimable linear function $C\beta$ is

$$C(X^\top X)^- X^\top y.$$

Uniqueness of the OLS Estimator of an Estimable $C\beta$

If $C\beta$ is estimable, then $C\hat{\beta}$ is the same for all solutions $\hat{\beta}$ to the normal equations.

In particular, the unique OLS estimator of $C\beta$ is

$$C\hat{\beta} = C(X^\top X)^- X^\top y = AX(X^\top X)^- X^\top y = AP_X y,$$

where $C = AX$.

Furthermore, if $C\beta$ is estimable, then $C\hat{\beta}$ is a **linear unbiased estimator** of $C\beta$.

The OLS estimator is linear because it is a linear function of y:

$$C\hat{\beta} = C(X^\top X)^- X^\top y = My,$$

where

$$M = C(X^\top X)^- X^\top.$$

The OLS estimator is unbiased because, for all $\beta \in \mathbb{R}^p$,

$$\begin{aligned} E(C\hat{\beta}) &= E\big(C(X^\top X)^- X^\top y\big) \\ &= C(X^\top X)^- X^\top E(y) \\ &= AX(X^\top X)^- X^\top X\beta \\ &= AP_X X\beta \\ &= AX\beta \\ &= C\beta. \end{aligned}$$

Gauss–Markov Model (GMM)

Suppose

$$y = X\beta + \varepsilon,$$

where

- $y \in \mathbb{R}^n$ is the response vector,
- X is an $n \times p$ matrix of known constants,
- $\beta \in \mathbb{R}^p$ is an unknown parameter vector, and
- ε is a vector of random errors satisfying

$$E(\varepsilon) = 0, \qquad \mathrm{Var}(\varepsilon) = \sigma^2 I,$$

for some unknown $\sigma^2 > 0$.

Gauss–Markov Theorem.
The OLS estimator of an estimable function $C\beta$ is the **Best Linear Unbiased Estimator (BLUE)** of $C\beta$, in the sense that it has the smallest variance among all linear unbiased estimators of $C\beta$.

Gauss–Markov Model with Normal Errors (GMMNE)

Suppose

$$y = X\beta + \varepsilon,$$

where

- $y \in \mathbb{R}^n$,
- X is an $n \times p$ matrix of known constants,
- $\beta \in \mathbb{R}^p$ is unknown, and
- $$\varepsilon \sim \mathcal{N}(0, \sigma^2 I).$$

Distribution of $C\hat{\beta}$ and $\hat{\sigma}^2$

In the GMMNE model, the distribution of $C\hat{\beta}$ is

$$C\hat{\beta} \sim \mathcal{N}\big(C\beta,\, \sigma^2\, C(X^\top X)^- C^\top\big)\,.$$

The distribution of $\hat{\sigma}^2$ is a scaled chi-square distribution:

$$\frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r},$$

equivalently,

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{n-r}\,\chi^2_{n-r}.$$

Moreover,

$$C\hat{\beta} \text{ and } \hat{\sigma}^2 \text{ are independent.}$$

F-Test

For $H_0 : C\beta = d$

To test

$$H_0 : C\beta = d,$$

use the statistic

$$F = \frac{(C\hat{\beta} - d)^\top \left[\mathrm{Var}(C\hat{\beta})\right]^{-1} (C\hat{\beta} - d)}{q}.$$

Since

$$\mathrm{Var}(C\hat{\beta}) = \sigma^2\, C(X^\top X)^- C^\top,$$

this becomes

$$F = \frac{(C\hat{\beta} - d)^\top \left[C(X^\top X)^- C^\top\right]^{-1} (C\hat{\beta} - d)/q}{\hat{\sigma}^2}.$$

Under H_0, F follows an F distribution with

$$q \quad \text{and} \quad n - r$$

degrees of freedom.

Under the alternative, F has a noncentral F distribution with noncentrality parameter

$$\theta = \frac{(C\beta - d)^\top \left[C(X^\top X)^- C^\top\right]^{-1} (C\beta - d)}{2\sigma^2}.$$

The non-negative non-centrality parameter

$$\frac{(C\beta - d)^\top \left[C(X^\top X)^- C^\top\right]^{-1} (C\beta - d)}{2\sigma^2}$$

is equal to zero if and only if $H_0 : C\beta = d$ is true.

If $H_0 : C\beta = d$ is true, the statistic F has a **central** F-distribution with

$$q \quad \text{and} \quad n - r$$

degrees of freedom, denoted $F_{q,n-r}$.

t-Test

For ($H_0 : c^\top \beta = d$) for Estimable $c^\top \beta$

Here, $c^\top$ is a row vector and d is a scalar ($q = 1$).

The test statistic is

$$t \equiv \frac{c^\top \hat{\beta} - d}{\sqrt{\widehat{\mathrm{Var}}(c^\top \hat{\beta})}} = \frac{c^\top \hat{\beta} - d}{\sqrt{\hat{\sigma}^2 \, c^\top (X^\top X)^- c}}.$$

The statistic t has a non-central t-distribution with non-centrality parameter

$$\frac{c^\top \beta - d}{\sqrt{\sigma^2 \, c^\top (X^\top X)^- c}},$$

and degrees of freedom

$$n - r.$$

The non-centrality parameter

$$\frac{c^\top \beta - d}{\sqrt{\sigma^2\, c^\top (X^\top X)^- c}}$$

is equal to zero if and only if $H_0 : c^\top \beta = d$ is true.

If $H_0 : c^\top \beta = d$ is true, the statistic t has a **central** t-distribution with

$$n - r$$

degrees of freedom, denoted t_{n-r}.

Confidence Interval

For Estimable $c^\top \beta$, a $100(1-\alpha)\%$ confidence interval for estimable $c^\top \beta$ is given by

$$c^\top \hat{\beta} \ \pm \ t_{n-r,\,1-\alpha/2} \sqrt{\hat{\sigma}^2\, c^\top (X^\top X)^- c}.$$

That is,

$$\text{estimate} \ \pm \ (\text{distribution quantile}) \times (\text{estimated standard error}).$$

Reduced vs. Full