2. Some Key Linear Models Results

A General Linear Model (GLM)

Suppose $y = X\beta + \epsilon$, 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 $E(\epsilon) = \mathbf{0}$.

This GLM says simply that y is a random vector satisfying $E(y) = X\beta$ for some $\beta \in \mathbb{R}^p$.

The distribution of y is left unspecified.

We know only that y is random and its mean is in the column space of X; i.e., $E(y) \in \mathcal{C}(X)$.

Ordinary Least Squares (OLS) Estimation of $E(\boldsymbol{y})$

Because we know $E(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X})$, a natural estimator of $E(\boldsymbol{y})$ is the *Ordinary Least Squares Estimator* (OLSE), which is the unique point in $\mathcal{C}(\boldsymbol{X})$ that is closest to $\boldsymbol{y}$ in terms of Euclidean distance.

The OLSE of $E(\boldsymbol{y})$ is given by $\hat{\boldsymbol{y}} \equiv \boldsymbol{P}_X \boldsymbol{y}$, where

$$\boldsymbol{P}_X = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}',$$

because $\boldsymbol{P}_X \boldsymbol{y} \in \mathcal{C}(\boldsymbol{X})$ and

$$||\boldsymbol{y} - \boldsymbol{P}_X \boldsymbol{y}||^2 < ||\boldsymbol{y} - \boldsymbol{z}||^2 \ \forall \ \boldsymbol{z} \in \mathcal{C}(\boldsymbol{X}) \setminus \{\boldsymbol{P}_X \boldsymbol{y}\}.$$

The Orthogonal-Projection Matrix

$P_X = X(X'X)^-X'$ is known as the *orthogonal-projection matrix* onto the column space of X and has the following properties:

- P_X is symmetric (i.e., $P_X = P_X'$).

- P_X is idempotent (i.e., $P_X P_X = P_X$).

- $P_X X = X$ and $X' P_X = X'$.

- $\text{rank}(X) = \text{rank}(P_X) = \text{tr}(P_X)$.

- $P_X = X(X'X)^-X'$ is the same matrix for all generalized inverses $(X'X)^-$ of $X'X$.

The OLSE of a Linear Function of $E(\boldsymbol{y})$

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

For any $q \times n$ matrix $\boldsymbol{A}$, the OLSE of $\boldsymbol{A}E(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta}$ is

$$\boldsymbol{A}[\text{OLSE of } E(\boldsymbol{y})] = \boldsymbol{A}\hat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{P}_X\boldsymbol{y} = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{y}.$$

The OLSE of an estimable function of $\boldsymbol{\beta}$

For any $q \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{A}E(\boldsymbol{y}) = \boldsymbol{AX\beta}$ is a linear function of $\boldsymbol{\beta}$ of the form $\boldsymbol{C\beta}$, where $\boldsymbol{C} = \boldsymbol{AX}$.

From the previous slide, we know

$\quad$ OLSE of $\boldsymbol{C\beta} = \boldsymbol{AX\beta} = \boldsymbol{A}E(\boldsymbol{y})$ is $\boldsymbol{AX}(\boldsymbol{X'X})^{-}\boldsymbol{X'y} = \boldsymbol{C}(\boldsymbol{X'X})^{-}\boldsymbol{X'y}$.

Now if $\boldsymbol{C}$ is any $q \times p$ matrix, we say that the linear function of $\boldsymbol{\beta}$ given by $\boldsymbol{C\beta}$ is *estimable* if and only if $\boldsymbol{C} = \boldsymbol{AX}$ for some matrix $q \times n$ matrix $\boldsymbol{A}$.

The OLSE of an estimable linear function $\boldsymbol{C\beta}$ is $\boldsymbol{C}(\boldsymbol{X'X})^{-}\boldsymbol{X'y}$.

The OLSE of Estimable Functions of $\boldsymbol{\beta}$

An equivalent definition for the OLSE of an estimable linear function of $\boldsymbol{\beta}$, given by $\boldsymbol{C\beta} = \boldsymbol{AX\beta}$, can be stated in terms of solutions to the

Normal Equations: $\qquad \boldsymbol{X'Xb} = \boldsymbol{X'y}$.

The OLSE of estimable $\boldsymbol{C\beta}$ is $\boldsymbol{C\hat{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is any solution for $\boldsymbol{b}$ in the Normal Equations.

Solutions to the Normal Equations

The Normal Equations

$$X'Xb = X'y$$

have $(X'X)^{-1}X'y$ as the unique solution for b if $\text{rank}(X) = p$.

The Normal Equations have infinitely many solutions for b if $\text{rank}(X) < p$.

$\hat{\beta} = (X'X)^- X'y$ is always a solution to the Normal Equations for any $(X'X)^-$, a generalized inverse of $X'X$.

Uniqueness of the OLSE 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 OLSE of $C\beta$ is

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

where $C = AX$.

The OLSE is a Linear Unbiased Estimator

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

The OLSE is a *linear estimator* because it's a linear function of y:

$$C\hat{\beta} = C(X'X)^- X'y = My, \text{ where } M = C(X'X)^- X'.$$

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

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

The Guass-Markov Model (GMM)

Suppose $y = X\beta + \epsilon$, 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 $E(\epsilon) = \mathbf{0}$ and $\mathrm{Var}(\epsilon) = \sigma^2 I$ for some unknown variance parameter $\sigma^2 \in \mathbb{R}^+$.

The GMM is a Special Case of the GLM

The GMM is a special case of the GLM presented previously.

We have added the assumption $\mathrm{Var}(\epsilon) = \sigma^2 I$; i.e., we assume the errors are uncorrelated and have constant variance.

All the results presented for the GLM hold for the GMM.

The Guass-Markov Theorem

For the GMM, we have an additional result provided by the *Gauss-Markov Theorem*:

The OLSE of an estimable function $C\beta$ is the

      Best Linear Unbiased Estimator (BLUE) of $C\beta$

in the sense that the OLSE $C\hat{\beta}$ has the smallest variance among all linear unbiased estimators of $C\beta$.

Unbiased Estimation of σ^2

An unbiased estimator of σ^2 under the GMM is given by

$$\hat{\sigma}^2 \equiv \frac{y'(I - P_X)y}{n - r}, \text{ where } r = \text{rank}(X).$$

Because $I - P_X = (I - P_X)(I - P_X) = (I - P_X)'(I - P_X)$,

$$
\begin{aligned}
y'(I - P_X)y &= y'(I - P_X)'(I - P_X)y = \{(I - P_X)y\}'\{(I - P_X)y\} \\
&= ||(I - P_X)y||^2 = ||y - P_X y||^2 \\
&= ||y - \hat{y}||^2 = \text{"Sum of Squared Errors" (SSE)}.
\end{aligned}
$$

Gauss-Markov Model with Normal Errors (GMMNE)

Suppose

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

where

- $\boldsymbol{y} \in \mathbb{R}^n$ is the response vector,

- $\boldsymbol{X}$ is an $n \times p$ matrix of known constants,

- $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown parameter vector, and

- $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ for some unknown variance parameter $\sigma^2 \in \mathbb{R}^+$.

The GMMNE is a special case of the GMM.

We have added the assumption ϵ is multivariate normal.

The GMMNE implies $y \sim N(X\beta, \sigma^2 I)$.

The GMMNE is useful for drawing statistical inferences regarding estimable $C\beta$.

Throughout the remainder of these slides we will assume

- the GMMNE model holds,

- C is a $q \times p$ matrix such that $C\beta$ is estimable,

- $\operatorname{rank}(C) = q$, and

- d is a known $q \times 1$ vector.

These assumptions imply $H_0 : C\beta = d$ is a *testable hypothesis*.

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

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

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

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

The F Test of $H_0 : C\beta = d$

The test statistic

$$
\begin{aligned}
F &\equiv (C\hat{\beta} - d)'[\widehat{\mathrm{Var}}(C\hat{\beta})]^{-1}(C\hat{\beta} - d)/q \\
&= (C\hat{\beta} - d)'[\hat{\sigma}^2 C(X'X)^- C']^{-1}(C\hat{\beta} - d)/q \\
&= \frac{(C\hat{\beta} - d)'[C(X'X)^- C']^{-1}(C\hat{\beta} - d)/q}{\hat{\sigma}^2}
\end{aligned}
$$

has a non-central F distribution with non-centrality parameter

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

and degrees of freedom q and $n - r$.

The F Test of $H_0 : C\beta = d$ (continued)

The non-negative non-centrality parameter

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

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 and $n - r$ degrees of freedom ($F_{q,n-r}$).

The F Test of $H_0 : C\beta = d$ (continued)

Thus, to test $H_0 : C\beta = d$, we compute the test statistic F and compare the observed value of F to the $F_{q,n-r}$ distribution.

If F is so large that it seems unlikely to have been a draw from the $F_{q,n-r}$ distribution, we reject H_0 and conclude $C\beta \neq d$.

The p-value of the test is the probability that a random variable with distribution $F_{q,n-r}$ matches or exceeds the observed value of the test statistic F.

The t Test of $H_0 : c'\beta = d$ for Estimable $c'\beta$

The test statistic

$$
\begin{aligned}
t &\equiv \frac{c'\hat{\beta} - d}{\sqrt{\widehat{\mathrm{Var}}(c'\hat{\beta})}} \\
&= \frac{c'\hat{\beta} - d}{\sqrt{\hat{\sigma}^2 c'(X'X)^- c}}
\end{aligned}
$$

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

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

and degrees of freedom $n - r$.

The t Test (continued)

The non-centrality parameter

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

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 t has a central t distribution with $n - r$ degrees of freedom (t_{n-r}).

The t Test (continued)

Thus, to test $H_0 : c'\beta = d$, we compute the test statistic t and compare the observed value of t to the t_{n-r} distribution.

If t is so far from zero that it seems unlikely to have been a draw from the t_{n-r} distribution, we reject H_0 and conclude $c'\beta \neq d$.

The p-value of the test is the probability that a random variable with distribution t_{n-r} would be as far or farther from 0 than the observed value of the t test statistic.

A $100(1 - \alpha)$% Confidence Interval for Estimable $\boldsymbol{c}'\boldsymbol{\beta}$

$$\left(\boldsymbol{c}'\hat{\boldsymbol{\beta}} - t_{n-r,1-\alpha/2}\sqrt{\hat{\sigma}^2 \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{c}}, \ \ \boldsymbol{c}'\hat{\boldsymbol{\beta}} + t_{n-r,1-\alpha/2}\sqrt{\hat{\sigma}^2 \boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{c}} \right)$$

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

estimate $\pm$ (distribution quantile) $\times$ (standard error of estimate)

Form of the t Statistic for Testing $H_0 : c'\beta = d$

$$t = \frac{\text{estimate} - d}{\text{standard error of estimate}} = \frac{\text{estimate} - d}{\sqrt{\widehat{\text{Var}}(\text{estimator})}}$$

$$t^2 = \frac{(\text{estimate} - d)^2}{\widehat{\text{Var}}(\text{estimator})}$$

$$= (\text{estimate} - d)\left[\widehat{\text{Var}}(\text{estimator})\right]^{-1}(\text{estimate} - d)/1$$

Revisiting the F Statistic for Testing $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$

$$
\begin{aligned}
F &= (\textbf{estimate} - \boldsymbol{d})' \left[\widehat{\mathrm{Var}}(\textbf{estimator}) \right]^{-1} (\textbf{estimate} - \boldsymbol{d})/q \\[2mm]
&= (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})' [\widehat{\mathrm{Var}}(\boldsymbol{C}\hat{\boldsymbol{\beta}})]^{-1} (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})/q \\[2mm]
&= (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})' [\hat{\sigma}^2 \boldsymbol{C}(\boldsymbol{X'X})^- \boldsymbol{C}']^{-1} (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})/q \\[2mm]
&= \frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})' [\boldsymbol{C}(\boldsymbol{X'X})^- \boldsymbol{C}']^{-1} (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})/q}{\hat{\sigma}^2}
\end{aligned}
$$