A Review of Some Key Linear Models Results

A General Linear Model (GLM)

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
- $\boldsymbol{\epsilon}$ is a vector of random "errors" satisfying $E(\boldsymbol{\epsilon}) = \mathbf{0}$.

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

The distribution of $\boldsymbol{y}$ is left unspecified.

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

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 $\hat{\boldsymbol{y}} \equiv \boldsymbol{P_X y}$, where

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

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

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

The Orthogonal Projection Matrix

$\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is known as the *orthogonal projection matrix* (a.k.a. the perpendicular projection operator) onto the column space of $\boldsymbol{X}$ and has the following properties:

- $\boldsymbol{P_X}$ is symmetric (i.e., $\boldsymbol{P_X} = \boldsymbol{P}'_{\boldsymbol{X}}$).
- $\boldsymbol{P_X}$ is idempotent (i.e., $\boldsymbol{P_X}\boldsymbol{P_X} = \boldsymbol{P_X}$).
- $\boldsymbol{P_X}\boldsymbol{X} = \boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{P_X} = \boldsymbol{X}'$.
- $\text{rank}(\boldsymbol{X}) = \text{rank}(\boldsymbol{P_X}) = \text{tr}(\boldsymbol{P_X})$.
- $\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'$ is the same matrix for all generalized inverses $(\boldsymbol{X}'\boldsymbol{X})^{-}$ of $\boldsymbol{X}'\boldsymbol{X}$.

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 $\boldsymbol{A}E(\boldsymbol{y}) = \boldsymbol{AX\beta}$ is

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

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

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

From the previous slide, we know

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

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

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

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

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

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

Solutions to the Normal Equations

The Normal Equations

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$$

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

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

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

Uniqueness of the OLSE of an Estimable $\boldsymbol{C\beta}$

If $\boldsymbol{C\beta}$ is estimable, then $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ is the same for all solutions $\hat{\boldsymbol{\beta}}$ to the Normal Equations.

In particular, the unique OLSE of $\boldsymbol{C\beta}$ is

$$\boldsymbol{C}\hat{\boldsymbol{\beta}} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y},$$

where $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{X}$.

The Guass-Markov Model (GMM)

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
- $\boldsymbol{\epsilon}$ is a vector of random "errors" satisfying $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$ for some unknown variance parameter $\sigma^2 \in \mathbb{R}^+$.

The Guass-Markov Theorem

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

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

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

For the GMM we have an additional result provided by the *Gauss-Markov Theorem*: The OLSE of an estimable function $\boldsymbol{C\beta}$ is the Best Linear Unbiased Estimator (BLUE) of $\boldsymbol{C\beta}$ in the sense that the OLSE $\boldsymbol{C}\hat{\boldsymbol{\beta}}$ has the smallest variance among all linear unbiased estimators of $\boldsymbol{C\beta}$.

Unbiased Estimation of σ^2

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

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

Note that

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

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

The GMMNE is a special case of the GMM.

We have added the assumption $\boldsymbol{\epsilon}$ is multivariate normal.

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

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

Throughout the remainder of these slides we will assume

- the GMMNE model holds,
- $\boldsymbol{C}$ is a $q \times p$ matrix such that $\boldsymbol{C\beta}$ is estimable,
- $\text{rank}(\boldsymbol{C}) = q$, and
- $\boldsymbol{d}$ is a known $q \times 1$ vector.

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

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

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

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

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

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

The test statistic

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

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

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

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

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

The non-negative non-centrality parameter

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

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

If $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ 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 $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$ (continued)

Thus, to test $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$, 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 $\boldsymbol{C\beta} \neq \boldsymbol{d}$.

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 $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = d$ for Estimable $\boldsymbol{c}'\boldsymbol{\beta}$

The test statistic

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

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

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

and degrees of freedom $n - r$.

The t Test (continued)

The non-centrality parameter

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

is equal to zero if and only if $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = d$ is true.

If $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = d$ 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 $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = d$, 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 $\boldsymbol{c}'\boldsymbol{\beta} \neq d$.

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 $H_0 : \boldsymbol{c}'\boldsymbol{\beta} = d$

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

$$\begin{aligned} t^2 &= \frac{(\text{estimate} - d)^2}{\widehat{\text{Var}}(\text{estimator})} \\ &= (\text{estimate} - d)\left[\widehat{\text{Var}}(\text{estimator})\right]^{-1}(\text{estimate} - d)/1 \end{aligned}$$

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

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