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,
- $oldsymbol{ heta} oldsymbol{eta} \in \mathbb{R}^p$ is an unknown parameter vector, and
- $oldsymbol{\epsilon}$ is a vector of random "errors" satisfying $E(oldsymbol{\epsilon})=oldsymbol{0}.$

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., $E(y) \in C(X)$.

Ordinary Least Squares (OLS) Estimation of E(y)

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

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

$$P_X = X(X'X)^-X',$$

because $P_X y \in C(X)$ and

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

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., $P_X P_X = P_X$).
- $P_XX = X$ and $X'P_X = X'$.
- $\operatorname{rank}(X) = \operatorname{rank}(P_X) = \operatorname{tr}(P_X).$
- $P_X = X(X'X)^-X'$ is the same matrix for all generalized inverses $(X'X)^-$ of X'X.

The OLSE of a Linear Function of E(y)

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

For any $q \times n$ matrix A, the OLSE of $AE(y) = AX\beta$ is

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

The OLSE of an estimable function of β

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

From the previous slide, we know

OLSE of
$$C\beta = AX\beta = AE(y)$$
 is $AX(X'X)^-X'y = C(X'X)^-X'y$.

Now 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 matrix $q \times n$ matrix A.

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

The OLSE of Estimable Functions of β

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

Normal Equations:
$$X'Xb = X'y$$
.

The OLSE of estimable $C\beta$ is $C\hat{\beta}$, where $\hat{\beta}$ is any solution for 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 rank(X) = p.

The Normal Equations have infinitely many solutions for b if 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

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

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:

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

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

$$E(C\hat{\boldsymbol{\beta}}) = E(C(X'X)^{-}X'y) = C(X'X)^{-}X'E(y) = AX(X'X)^{-}X'E(y)$$
$$= AP_{X}E(y) = AP_{X}X\boldsymbol{\beta} = AX\boldsymbol{\beta} = C\boldsymbol{\beta}.$$

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 $\mathrm{Var}(\epsilon) = \sigma^2 \mathbf{I}$ 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 $Var(\epsilon) = \sigma^2 I$; 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 $\hat{C\beta}$ has the smallest variance among all linear unbiased estimators of $\hat{C\beta}$.

Unbiased Estimation of σ^2

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

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

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

$$y'(I - P_X)y = y'(I - P_X)'(I - P_X)y = \{(I - P_X)y\}'\{(I - P_X)y\}$$

$$= ||(I - P_X)y||^2 = ||y - P_Xy||^2$$

$$= ||y - \hat{y}||^2 = \text{"Sum of Squared Errors" (SSE)}.$$

Gauss-Markov Model with Normal Errors (GMMNE)

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,
- $oldsymbol{eta} \in \mathbb{R}^p$ is an unknown parameter vector, and
- $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ 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 $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$.

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,
- $rank(\mathbf{C}) = q$, and
- d is a known $q \times 1$ vector.

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

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

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

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

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

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

The test statistic

$$F \equiv (C\hat{\boldsymbol{\beta}} - \boldsymbol{d})'[\widehat{\text{Var}}(C\hat{\boldsymbol{\beta}})]^{-1}(C\hat{\boldsymbol{\beta}} - \boldsymbol{d})/q$$

$$= (C\hat{\boldsymbol{\beta}} - \boldsymbol{d})'[\hat{\sigma}^2 \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^- \boldsymbol{C}']^{-1}(C\hat{\boldsymbol{\beta}} - \boldsymbol{d})/q$$

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

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

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

and degrees of freedom q and n - r.

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

The non-negative non-centrality parameter

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

is equal to zero if and only if H_0 : $C\beta = d$ 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 $F_{a,n-r}$ 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

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

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 : $c'\beta = d$ 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 $c'\beta$

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

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

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

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

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

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

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

Revisiting the *F* Statistic for Testing H_0 : $C\beta = d$

$$F = (\mathbf{estimate} - \mathbf{d})' \Big[\widehat{\mathbf{Var}} (\mathbf{estimator}) \Big]^{-1} (\mathbf{estimate} - \mathbf{d})/q$$

$$= (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})' [\widehat{\mathbf{Var}} (\mathbf{C}\hat{\boldsymbol{\beta}})]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})/q$$

$$= (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})' [\hat{\sigma}^2 \mathbf{C} (\mathbf{X}'\mathbf{X})^- \mathbf{C}']^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})/q$$

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