# Chapter 3 Testing

Hakbae Lee

The Department of Statistics and Data Science, Yonsei University

# More About Models: Two approaches for linear model

- Parameter-free approach

$$\mathbf{Y} = \mathrm{E}(\mathbf{Y}) + \mathbf{Y} - \mathrm{E}(\mathbf{Y}) = \mu + \boldsymbol{\epsilon}$$

  where $\mathrm{E}(\mathbf{Y}) = \mu$ and $\boldsymbol{\epsilon} = \mathbf{Y} - \mathrm{E}(\mathbf{Y})$.

- Parameter approach

$$\mathbf{Y} = \mathrm{E}(\mathbf{Y}) + \mathbf{Y} - \mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

  where $\mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$.

- $$\mathrm{E}(\boldsymbol{\epsilon}) = 0,\ \mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2 I \quad \text{Ordinary Least Square(OLS)}$$

or

$$\mathrm{E}(\boldsymbol{\epsilon}) = 0,\ \mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \Sigma \quad \text{Generalized Least Square(GLS)}$$

# More About Models

- Consider

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad \mathrm{E}(\epsilon) = 0, \quad \mathrm{Cov}(\epsilon) = \sigma^2 I$$

- $\mathcal{C}(\mathbf{X}) =$ Estimation space
- $M = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T =$ orthogonal projection onto $\mathcal{C}(\mathbf{X})$
- $\mathrm{E}(\mathbf{Y}) = \mathbf{X}\beta \in \mathcal{C}(\mathbf{X})$, $\mathrm{Cov}(\mathbf{Y}) = \sigma^2 I$
- $\mathcal{C}(\mathbf{X})^\perp =$ Error space
- $\mathbf{I} - M = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T =$ orthogonal projection onto $\mathcal{C}(\mathbf{X})^\perp$
- $\epsilon \in \mathcal{C}(\mathbf{X})^T$, $\mathrm{Cov}(\epsilon) = \sigma^2 I$
- Any two linear models with the same estimation space are the same model.

# More About Models

- *One-Way ANOVA*

$$\begin{aligned} y_{ij} &= \mu_i + \epsilon_{ij} = \mu + \alpha_i + \epsilon_{ij}, \text{ with } \mu_i = \mathrm{E}(y_{ij}) = \mu + \alpha_i \\ \bar{\mu} &= \mu + \bar{\alpha}_+ \\ \mu_1 - \mu_2 &= \alpha_1 - \alpha_2 \end{aligned}$$

  The parameters in the two models are different, but they are related.

- *Simple Linear Regression*

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\ &= \gamma_0 + \gamma_1 (x_i - \bar{x}) + \epsilon_i \end{aligned}$$

where

$$\mathrm{E}(y_i) = \beta_0 + \beta_1 x_i = \gamma_0 + \gamma_1 (x_i - \bar{x})$$

# More About Models

- $$\mathcal{C}(\mathbf{X}_1) = \mathcal{C}(\mathbf{X}_1) \Rightarrow \mathbf{X}_1 = \mathbf{X}_2 T$$

  so that

  $$\mathbf{X}_1 \boldsymbol{\beta}_1 = \mathbf{X}_2 T \boldsymbol{\beta}_1 = \mathbf{X}_2 \boldsymbol{\beta}_2$$

  which implies that

  $$\boldsymbol{\beta}_2 = T\boldsymbol{\beta}_1 + \nu \quad \text{for } \nu \in \mathcal{C}(\mathbf{X}_2^T)^\perp$$

- NOTE: A unique parameterization for $\mathbf{X}_j, j = 1, 2$ occurs if and only if $\mathbf{X}_j^T \mathbf{X}_j$ is nonsingular.

- Exercise: Show that a unique parameterization for $\mathbf{X}_j, j = 1, 2$ means $\mathcal{C}(\mathbf{X}_2^T)^\perp = \{0\}$.

# Testing Models

- Consider

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad \epsilon \sim N(0, \mathbf{I}_n)$$

Let's partition $\mathbf{X}$ into $\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1)$: $\mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$

$$\mathbf{Y} = \mathbf{X}_0\beta_0 + \mathbf{X}_1\beta_1 + \epsilon \quad : \text{Full Model(FM)}$$

$$\mathbf{Y} = \mathbf{X}_0\gamma + \epsilon \quad : \text{Reduced Model(RM)}$$

- Hypothesis testing procedure can be described as

$$H_0 : \text{Reduced Model(RM)} \quad H_1 : \text{Full Model(FM)}$$

- Example 3.2.0: pp. 52–54

- Let $M$ and $M_0$ be the orthogonal projection onto $\mathcal{C}(\mathbf{X})$ and $\mathcal{C}(\mathbf{X}_0)$ respectively.
- Note that with $\mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$, $M - M_0$ is the orthogonal projection onto the orthogonal complement of $\mathcal{C}(\mathbf{X}_0)$ with respect to $\mathcal{C}(\mathbf{X})$, that is,

$$\mathcal{C}(X_0)^{\perp}_{\mathcal{C}(\mathbf{X})} = \mathcal{C}(M - M_0) = \mathcal{C}(M \cap M_0^{\perp}) \qquad \text{check!}$$

  and

$$\begin{aligned} \hat{\mu} &= \hat{\mathrm{E}}(\mathbf{Y}) = M\mathbf{Y} \qquad \text{under FM} \\ \hat{\mu}_0 &= \hat{\mathrm{E}}(\mathbf{Y}) = M_0\mathbf{Y} \qquad \text{under RM} \end{aligned}$$

- If RM is true, then $M\mathbf{Y} - M_0\mathbf{Y} = (M - M_0)\mathbf{Y}$ should be reasonably small.
- Note that $\mathrm{E}(M - M_0)\mathbf{Y} = 0$

# Testing Models

The decision about whether RM is appropriate hinges on deciding whether the vector $(M - M_0)\mathbf{Y}$ is large.

An obvious measure of the size of $(M - M_0)\mathbf{Y}$ is

$$[(M - M_0)\mathbf{Y}]^T[(M - M_0)\mathbf{Y}] = \mathbf{Y}^T(M - M_0)\mathbf{Y}.$$

A reasonable measure of the size of $(M - M_0)\mathbf{Y}$ is given by $\mathbf{Y}^T(M - M_0)\mathbf{Y}/r(M - M_0)$

Note that

$$\mathrm{E}\left[\frac{\mathbf{Y}^T(M - M_0)\mathbf{Y}}{r(M - M_0)}\right] = \sigma^2 + \frac{\beta^T\mathbf{X}^T(M - M_0)\mathbf{X}\beta}{r(M - M_0)}$$

**Theorem 3.2.1.** Consider $\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad \epsilon \sim N(0, \mathbf{I}_n)$ with $\mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_0\beta_0 + \mathbf{X}_1\beta_1 + \epsilon && : \text{Full Model(FM)} \\ \mathbf{Y} &= \mathbf{X}_0\gamma + \epsilon && : \text{Reduced Model(RM)} \end{aligned}$$

(i) Under the Full Model(FM),

$$\frac{\mathbf{Y}^T(M - M_0)\mathbf{Y}/r(M - M_0)}{\mathbf{Y}^T(\mathbf{I} - M)\mathbf{Y}/r(\mathbf{I} - M)} \sim F\left(df_1, df_2, \frac{\beta^T\mathbf{X}^T(M - M_0)\mathbf{X}\beta}{2\sigma^2}\right)$$

where $df_1 = r(M - M_0)$, $df_2 = r(\mathbf{I} - M)$

(ii) Under the Reduced Model(RM),

$$\frac{\mathbf{Y}^T(M - M_0)\mathbf{Y}/r(M - M_0)}{\mathbf{Y}^T(\mathbf{I} - M)\mathbf{Y}/r(\mathbf{I} - M)} \sim F(df_1, df_2, 0)$$

where $df_1 = r(M - M_0)$, $df_2 = r(\mathbf{I} - M)$

**Note:**

$$
\begin{aligned}
M - M_0 &= (I - M_0) - (I - M) \\
\mathbf{Y}^T(M - M_0)\mathbf{Y} &= \mathbf{Y}^T(I - M_0)\mathbf{Y} - \mathbf{Y}^T(I - M)\mathbf{Y} \\
&= \mathrm{SSE}_{\mathrm{RM}} - \mathrm{SSE}_{\mathrm{FM}}
\end{aligned}
$$

Example 3.2.2.; pp. 58–59

# A Generalized Test Procedure

Assume that

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \text{ is correct.}$$

Want to test the adequacy of a model

$$\mathbf{Y} = \mathbf{X}_0\gamma + \mathbf{X}b + \epsilon$$

where $\mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$ and some known vector $\mathbf{X}b =$ offset

Example 3.2.3.; Multiple Regression

$$\mathbf{Y} = \beta_0 J + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

To test $H_0 : \beta_2 = \beta_3 + 5, \beta_1 = 0$,....

# A Generalized Test Procedure

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}\beta + \epsilon \quad \text{(FM)} \\ \mathbf{Y}^* &\equiv \mathbf{Y} - \mathbf{X}b = \mathbf{X}\beta - \mathbf{X}b + \epsilon = \mathbf{X}\beta^* + \epsilon \quad \text{(FM)}\end{aligned}$$

where $\beta^* = \beta - b$ and let's consider

$$\begin{aligned}\mathbf{Y} &= \mathbf{X}_0\gamma + \mathbf{X}b + \epsilon \quad \text{(RM)} \\ \mathbf{Y}^* &= \mathbf{Y} - \mathbf{X}b = \mathbf{X}\gamma + \epsilon \quad \text{(RM)}\end{aligned}$$

In addition,

$$\frac{\mathbf{Y}^{*T}(M - M_0)\mathbf{Y}^*/r(M - M_0)}{\mathbf{Y}^{*T}(I - M)\mathbf{Y}^*/r(I - M)} \sim F(r(M - M_0), r(I - M), \delta^2)$$

where the noncentrality parameter $\delta^2$ is

$$\delta^2 = \frac{1}{2\sigma^2}\beta^{*T}\mathbf{X}^T(M - M_0)\mathbf{X}\beta^*$$

$$0 = \beta^{*T}\mathbf{X}^T(M - M_0)\mathbf{X}\beta^*$$

if and only if

$$0 = (M - M_0)\mathbf{X}\beta^* \qquad \text{why?}$$

if and only if

$$\mathbf{X}\beta = M_0(\mathbf{X}\beta - \mathbf{X}b) + \mathbf{X}b \qquad \text{why?}$$

which holds if

$$\gamma = (\mathbf{X}_0^T\mathbf{X}_0)^{-}\mathbf{X}_0(\mathbf{X}\beta - \mathbf{X}b) = (\mathbf{X}_0^T\mathbf{X}_0)^{-}\mathbf{X}_0\mathbf{X}\beta^*$$

Furthermore,

$$\mathbf{Y}^{*T}(M - M_0)\mathbf{Y}^* = \mathbf{Y}^{*T}(I - M_0)\mathbf{Y}^* - \mathbf{Y}^{*T}(I - M)\mathbf{Y}^*$$

and

$$\mathbf{Y}^{*T}(I - M)\mathbf{Y}^* = \mathbf{Y}^T(I - M)\mathbf{Y}$$

Example 3.2.3.

# Testing Linear Parametric Functions

$$H_0 : \mathbf{Y} = \mathbf{X}\beta + \epsilon \quad \text{and} \quad \Lambda^T\beta = 0 \qquad (1)$$

$$\begin{aligned}
\Lambda^T\beta = 0 \quad &\iff \quad \beta \in \mathcal{N}(\Lambda^T) = \mathcal{C}(X)^\perp \\
&\iff \quad \beta \perp \mathcal{C}(\Lambda) \\
&\iff \quad \beta \perp \mathcal{C}(\Gamma) \quad \text{if } \exists\, \Gamma \text{ so that } \mathcal{C}(\Gamma) = \mathcal{C}(\Lambda) \\
&\iff \quad \beta \perp \mathcal{C}(U) \quad \text{if } \exists\, U \text{ so that } \mathcal{C}(U) = \mathcal{C}(\Lambda)^\perp \\
&\iff \quad \beta = U\gamma \quad \text{for some } \gamma \qquad (2)
\end{aligned}$$

Thus, letting $\mathbf{X}_0 = \mathbf{X}U$, (in general, $\mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$)

$$\mathbf{Y} \quad = \quad \mathbf{X}\beta + \epsilon = \mathbf{X}U\gamma + \epsilon = \mathbf{X}_0\gamma + \epsilon \qquad (3)$$

Suppose $\mathcal{C}(\mathbf{X}_0) = \mathcal{C}(\mathbf{X})$. Then there is nothing to test and $\Lambda^T\beta = 0$ involves only arbitrary side conditions that do not affect the model.

EXAMPLE 3.3.1. pp. 62–64

$$\text{Estimable } \Lambda^T\beta \quad \Longleftrightarrow \quad \Lambda = \mathbf{X}^T P \quad \text{for some } P$$

**Remark:**

$$\mathcal{C}(MP) \equiv \mathcal{C}(M - M_0) = \mathcal{C}(\mathbf{X} - \mathbf{X}_0) = \mathcal{C}(\mathbf{X}) \cap \mathcal{C}(\mathbf{X}_0)^\perp = \mathcal{C}(\mathbf{X}_0)^\perp_{\mathcal{C}(\mathbf{X})}$$

Thus, its distribution for testing $H_0 : \Lambda^T\beta = 0$ is given by

$$\frac{\mathbf{Y}^T M_{MP}\mathbf{Y}/r(M_{MP})}{\mathbf{Y}^T(I - M)\mathbf{Y}/r(I - M)} \sim F(r(M_{MP}), r(I - M), \delta^2) \qquad (5)$$

where $\delta^2 = \beta^T\mathbf{X}^T M_{MP}\mathbf{X}\beta$

**Proposition 3.3.2**

$$\mathcal{C}(M - M_0) = \mathcal{C}(\mathbf{X}_0)^\perp_{\mathcal{C}(\mathbf{X})} \equiv \mathcal{C}(\mathbf{X}U)^\perp_{\mathcal{C}(\mathbf{X})} = \mathcal{C}(MP)$$

$$
\begin{aligned}
& H_0 : \mathbf{Y} = \mathbf{X}\beta + \epsilon \quad \text{and} \quad \Lambda^T \beta = 0 \\
\iff \quad & H_0 : \mathbf{Y} = \mathbf{X}\beta + \epsilon \quad \text{and} \quad P^T \mathbf{X}\beta = 0 \\
\iff \quad & H_0 : \mathbf{Y} = \mathbf{X}\beta + \epsilon \quad \text{and} \quad P^T M\mathbf{X}\beta = 0 \quad (M\mathbf{X} = \mathbf{X}) \\
\iff \quad & \mathrm{E}(\mathbf{Y}) \in \mathcal{C}(\mathbf{X}) \quad \text{and} \quad \mathrm{E}(\mathbf{Y}) \perp \mathcal{C}(MP) \\
\iff \quad & \mathrm{E}(\mathbf{Y}) \in \mathcal{C}(\mathbf{X}) \cap \mathcal{C}(MP)^{\perp}
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathcal{C}(\mathbf{X}_0) = \mathcal{C}(\mathbf{X}) \cap \mathcal{C}(MP)^{\perp} = \mathcal{C}(MP)^{\perp}_{\mathcal{C}(\mathbf{X})} \\
\Rightarrow \quad & \mathcal{C}(\mathbf{X}_0)^{\perp}_{\mathcal{C}(\mathbf{X})} = \mathcal{C}(MP) \\
\iff \quad & \mathbf{X}_0 = (I - M_{MP})\mathbf{X}
\end{aligned}
$$

**Theorem 3.3.3**

$$
\mathcal{C}[(I - M_{MP})\mathbf{X}] = \mathcal{C}(\mathbf{X}) \cap \mathcal{C}(MP)^{\perp} = \mathcal{C}(\mathbf{X}) \cap \mathcal{C}(P)^{\perp}
$$

EXAMPLE 3.3.4.: pp.66–67

$$
\begin{aligned}
\Lambda^T\beta \quad & \text{is} \quad \text{estimable, i.e., } \Lambda = \mathbf{X}^T P \\
\mathcal{C}(\Lambda) \quad & = \quad \mathcal{C}(\mathbf{X}^T P) = \mathcal{C}(MP)
\end{aligned}
$$

and $\mathbf{X}\hat{\beta} = M\mathbf{Y}$, and $\Lambda^T\hat{\beta} = P^T\mathbf{X}\hat{\beta} = P^T M\mathbf{Y}$

$$
\begin{aligned}
\mathbf{Y}^T M_{MP}\mathbf{Y} \quad & = \quad \mathbf{Y}^T M(P^T MP)^- MP\mathbf{Y} \\
& = \quad \hat{\beta}^T \Lambda[P^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T P]^-\Lambda^T\hat{\beta} \\
& = \quad \hat{\beta}^T \Lambda[\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda]^-\Lambda^T\hat{\beta}
\end{aligned}
$$

Thus,

$$
(5) \quad = \quad \frac{\hat{\beta}^T \Lambda[\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda]^-\Lambda^T\hat{\beta}/r(\Lambda)}{MSE} \sim F(r(MP), r(I-M), \delta^2)
$$

where

$$
\begin{aligned}
\delta^2 \quad & = \quad \frac{\beta^T \Lambda[\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda]^-\Lambda^T\beta}{2\sigma^2} \\
\mathrm{Cov}(\Lambda^T\hat{\beta}) \quad & = \quad \sigma^2\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda
\end{aligned}
$$

# Testing Linear Parametric Functions

For $H_0 : \lambda^T \boldsymbol{\beta} = 0, \ \lambda \in \mathbf{R}^p$,

$$\begin{aligned} \mathbf{Y}^T M_{MP} \mathbf{Y} &= \hat{\boldsymbol{\beta}}^T \lambda [\lambda^T (\mathbf{X}^T \mathbf{X})^- \lambda]^- \lambda^T \hat{\boldsymbol{\beta}} \\ &= \frac{(\lambda^T \hat{\boldsymbol{\beta}})^2}{\lambda^T (\mathbf{X}^T \mathbf{X})^- \lambda} \end{aligned}$$

and, under $H_0 : \lambda^T \boldsymbol{\beta} = 0$

$$F = (5) \quad = \quad \frac{(\lambda^T \hat{\boldsymbol{\beta}})^2}{MSE[\lambda^T (\mathbf{X}^T \mathbf{X})^- \lambda]} \sim F(1, r(I - M))$$

**Definition 3.3.5.** The condition $\mathrm{E}(\mathbf{Y}) \perp \mathcal{C}(MP)$ is called the constraint by $\Lambda^T \boldsymbol{\beta} = 0$ where $\Lambda = \mathbf{X}^T P$: $\mathcal{C}(MP) =$ the constraint by $\Lambda^T \boldsymbol{\beta} = 0$.

**Do Exercise 3.5:** Show that a necessary and sufficient condition for $\rho_1^T \mathbf{X} \boldsymbol{\beta} = 0$ and $\rho_2^T \mathbf{X} \boldsymbol{\beta} = 0$ to determine orthogonal constraints on the model is that $\rho_1^T \mathbf{X} \rho_2 = 0$.

# Theoretical Complements

- Consider testing $\Lambda^T\beta = 0$ when $\Lambda^T\beta$ is NOT estimable

- Let $\Lambda_0^T\beta$ be estimable part of $\Lambda^T\beta$.

- $\Lambda_0$ is chosen so that

$$\mathcal{C}(\Lambda_0) = \mathcal{C}(\Lambda) \cap \mathcal{C}(\mathbf{X}^T)$$

which means that $\Lambda^T\beta = 0$ implies that $\Lambda_0^T\beta = 0$ but $\Lambda_0^T\beta$ is estimable because

$$\mathcal{C}(\Lambda_0) \subset \mathcal{C}(\mathbf{X}^T).$$

# Theoretical Complements

**Theorem 3.3.6.** Let $\mathcal{C}(\Lambda_0) = \mathcal{C}(\Lambda) \cap \mathcal{C}(\mathbf{X}^T)$ and $\mathcal{C}(U_0) = \mathcal{C}(\Lambda_0)^\perp$. Then $\mathcal{C}(\mathbf{X}U) = \mathcal{C}(\mathbf{X}U_0)$. Thus $\Lambda^T\beta = 0$ and $\Lambda_0^T\beta = 0$ induce the same reduced model.

**Proposition 3.3.7.** Let $\Lambda_0^T\beta$ be estimable and $\Lambda \neq 0$. Then

$$\Lambda^T\beta = 0 \Longrightarrow \mathcal{C}(\mathbf{X}U) \neq \mathcal{C}(\mathbf{X}).$$

**Corollary 3.3.8.**

$$\mathcal{C}(\Lambda_0) = \mathcal{C}(\Lambda) \cap \mathcal{C}(\mathbf{X}^T) = \{0\} \Longleftrightarrow \mathcal{C}(\mathbf{X}U) = \mathcal{C}(\mathbf{X}).$$

# A Generalized Test Procedure

Consider $H_0 : \Lambda^T\beta = d$, where $d \in \mathcal{C}(\mathbf{X}^T)$, which is solvable. Let $b$ so that $\Lambda^T b = d$. Then

$$\begin{aligned}
\Lambda^T\beta = \Lambda^T b = d \Longleftrightarrow \Lambda^T(\beta - b) = 0 \\
\Longleftrightarrow \quad & (\beta - b) \perp \mathcal{C}(\Lambda) \\
\Longleftrightarrow \quad & (\beta - b) \in \mathcal{C}(U) \quad \text{where} \quad \mathcal{C}(U) = \mathcal{C}(\Lambda)^\perp \\
\Longleftrightarrow \quad & \beta - b = U\gamma \quad \text{for some } \gamma \\
\Longleftrightarrow \quad & \mathbf{X}\beta - \mathbf{X}b = \mathbf{X}U\gamma \quad \text{i.e. } \mathbf{X}\beta = \mathbf{X}U\gamma + \mathbf{X}b
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\beta + \epsilon = \mathbf{X}U\gamma + \mathbf{X}b + \epsilon \\
&= \mathbf{X}_0\gamma + \mathbf{X}b + \epsilon \quad \text{where } \mathbf{X}_0 = \mathbf{X}U \qquad (6)
\end{aligned}$$

# A Generalized Test Procedure

If $\Lambda = \mathbf{X}^T P$, then $\mathcal{C}(\mathbf{X}_0)^{\perp}_{\mathcal{C}(\mathbf{X})} = \mathcal{C}(MP)$ and its test statistic is

$$\begin{aligned} F &= \frac{(\mathbf{Y} - \mathbf{X}b)^T M_{MP}(\mathbf{Y} - \mathbf{X}b)/r(M_{MP})}{(\mathbf{Y} - \mathbf{X}b)^T(I - M)(\mathbf{Y} - \mathbf{X}b)/r(I - M)} \\ &= \frac{(\Lambda^T\hat{\beta} - d)^T[\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda]^-(\Lambda^T\hat{\beta} - d)/r(\Lambda)}{MSE} \sim F(?, ?, ?) \end{aligned}$$

**Remark:** If $\Lambda^T\beta = d$, the same reduced model results if we take $\Lambda^T\beta = d_0$ where $d_0 = d + \Lambda^T\nu$ and $\nu \perp \mathcal{C}(\mathbf{X}^T)$. Note that, in this construction, if $\Lambda^T\beta = d$ is estimable, $d_0 = d$ for any $\nu$.

EXAMPLE 3.3.9.: pp.71–72

EXAMPLE 3.4.1.: pp.75

# Testing Single Degrees of Freedom in a Given Subspace

$$\mathbf{Y} = \mathbf{X}_0\gamma + \epsilon \text{ :RM} \quad \text{vs} \quad \mathbf{Y} = \mathbf{X}\beta + \epsilon \text{ :FM} \quad \text{With } \mathcal{C}(\mathbf{X}_0) \subset \mathcal{C}(\mathbf{X})$$

Let $M_* = M - M_0$. Consider $H_0 : \Lambda^T\beta = 0$.

If $\Lambda = \mathbf{X}^T P$, i.e., $\Lambda \in \mathcal{C}(X^T)$, then $M_* = M_{MP}$.

**Proposition 3.3.2** Since $MM_* = M_*$,

$$\mathcal{C}(M - M_0) = \mathcal{C}(\mathbf{X}_0)^{\perp}_{\mathcal{C}(\mathbf{X})} \equiv \mathcal{C}(\mathbf{X}U)^{\perp}_{\mathcal{C}(\mathbf{X})} = \mathcal{C}(MP)$$

$$\Rightarrow \quad M\rho \in \mathcal{C}(M_*) \Rightarrow M\rho = M_*M\rho = M_*\rho \Rightarrow \rho^T\hat{\beta} = \rho^T M_*\mathbf{Y} = \rho^T M\mathbf{Y}$$

Thus the test statistic for $H_0 : \Lambda^T\beta = 0$ is

$$\frac{\mathbf{Y}^T M_*\rho(\rho^T M_*\rho)^{-1}\rho^T M_*\mathbf{Y}}{MSE} = \frac{(\rho^T M_*\mathbf{Y})^2/\rho^T M_*\rho}{MSE}$$

# Breaking SS into Independent Components

Consider $\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1)$. Set

$$
\begin{aligned}
\text{SSR}(\mathbf{X}_1|\mathbf{X}_0) &\equiv \mathbf{Y}^T(M - M_0)\mathbf{Y} \\
&= \text{Sum of Squares for regression } \mathbf{X}_1 \text{ after } \mathbf{X}_0 \\
\text{SSR}(\mathbf{X}) &\equiv \mathbf{Y}^T M\mathbf{Y} \\
\text{SSR}(\mathbf{X}_0) &\equiv \mathbf{Y}^T M_0\mathbf{Y} \\
\text{SSR}(\mathbf{X}) &= \text{SSR}(\mathbf{X}_0) + \text{SSR}(\mathbf{X}_1|\mathbf{X}_0)
\end{aligned}
$$

**NOTE:** If $\epsilon \sim N(0, \sigma\mathbf{I})$, then

$$\text{SSR}(\mathbf{X}_0) \quad \perp\!\!\!\perp \quad \text{SSR}(\mathbf{X}_1|\mathbf{X}_0)$$

# General Theory

Let $M$ and $M_*$ be the orthogonal projection operator into $\mathcal{C}(\mathbf{X})$ and $\mathcal{C}(\mathbf{X}_*)$ respectively. Then, with $\mathcal{C}(\mathbf{X}_*) \subset \mathcal{C}(\mathbf{X})$, $M_*$ defines a test statistic

$$\frac{\mathbf{Y}^T M_* \mathbf{Y} / r(M_*)}{\mathbf{Y}^T (\mathbf{I} - M) \mathbf{Y} / r(\mathbf{I} - M)} \quad \text{for RM: } \mathbf{Y} = \mathbf{X}_* \gamma + \epsilon$$

$$\begin{aligned}
\mathcal{C}(M - M_*) &= \text{Estimation space under } H_0 \\
\mathcal{C}(M_*) &= \text{Test space under } H_0 \\
\mathcal{C}(\mathbf{I} - (M - M_*)) &= \text{Error space under } H_0
\end{aligned}$$

and $\mathbf{I} - (M - M_*) = (\mathbf{I} - M) + M_*$

Using Gram-Schmidt procedure, let's construct $M_*$ so that

$$M_* \quad = \quad RR^T = \sum_{i=1}^{r} R_i R_i^T = \sum_{i=1}^{r} M_i \quad \text{where } R = (R_1, \ldots, R_r)$$

and $M_i M_j = 0$ for $i \neq j$. By **Theorem 1.3.7**,

$$\mathbf{Y}^T M_i \mathbf{Y} \perp\!\!\!\perp \mathbf{Y}^T M_j \mathbf{Y} \Longleftrightarrow M_i M_j = 0$$

Next, $\mathbf{Y}^T M \mathbf{Y} = \sum_{i=1}^{r} \mathbf{Y}^T M_i \mathbf{Y}$, when $r(M_i) = 1$,

$$\frac{\mathbf{Y}^T M_i \mathbf{Y}/r(M_i)}{\mathbf{Y}^T(\mathbf{I} - M)\mathbf{Y}/r(\mathbf{I} - M)} \sim F\left(1, r(\mathbf{I} - M), \frac{1}{2\sigma^2}\beta^T \mathbf{X}^T M_i \mathbf{X}\beta\right)$$

# General Theory

$$
\begin{aligned}
0 &= \beta^T \mathbf{X}^T M_* \mathbf{X}\beta = \sum_{i=1}^{r} \beta^T \mathbf{X}^T M_i \mathbf{X}\beta \\
&\iff \beta^T \mathbf{X}^T M_i \mathbf{X}\beta = 0 \qquad \text{for all } i \\
&\iff R_i^T \mathbf{X}\beta = 0 \qquad \text{for all } i \\
&\iff H_0 \text{ is true}
\end{aligned}
$$

EXAMPLE 3.6.1.: Balanced design; pp.79–80

EXAMPLE 3.6.2.: Unbalanced design;pp.80–81

$$
\begin{aligned}
y_{ijk} &= \mu + \alpha_i + \eta_j + \epsilon_{ijk} : \text{FM} \\
M &= M_\mu + M_\alpha + M_\eta \\
y_{ijk} &= \mu + \alpha_i + \epsilon_{ijk} : \text{RM} \\
\mathbf{Y}^T(M - M_0)\mathbf{Y} &\equiv R(\eta|\alpha, \mu) \\
&= \text{Reduction in } SSE \text{ due to fitting } \eta_j\text{'s after } \mu \text{ and } \alpha_i\text{'s}
\end{aligned}
$$

Next,

$$
\begin{aligned}
y_{ijk} &= \mu + \alpha_i + \epsilon_{ijk} : \text{FM} \\
y_{ijk} &= \mu + \epsilon_{ijk} : \text{RM} \\
\mathbf{Y}^T(M_0 - M_J)\mathbf{Y} &= R(\alpha|\mu)
\end{aligned}
$$

# Two-Way ANOVA

$$\begin{aligned} y_{ijk} &= \mu + \alpha_i + \epsilon_{ijk} : \text{FM} \\ y_{ijk} &= \mu + \epsilon_{ijk} : \text{RM} \\ \mathbf{Y}^T(M - M_J)\mathbf{Y} &= R(\alpha, \eta|\mu) \\ &= R(\eta|\mu, \alpha) + R(\alpha|\mu) \end{aligned}$$

In general,

$$\begin{aligned} R(\eta|\alpha, \mu) &\neq R(\eta|\mu) \\ R(\alpha|\eta, \mu) &\neq R(\alpha|\mu) \end{aligned}$$

In paricular, for balanced design, if $\mathcal{C}(\mathbf{X}_\alpha) \perp \mathcal{C}(\mathbf{X}_\eta)$,

$$\begin{aligned} R(\eta|\alpha, \mu) &= R(\eta|\mu) \\ R(\alpha|\eta, \mu) &= R(\alpha|\mu) \end{aligned}$$

**Proposition 3.6.3.**

$$R(\eta|\alpha,\mu) = R(\eta|\mu) \Longleftrightarrow \mathcal{C}(M_1 - M_J) \perp \mathcal{C}(M_0 - M_J)$$

that is,

$$M_1 - M_J = M - M_0 \Longleftrightarrow (M_1 - M_J)(M_0 - M_J) = 0$$

where

$$\begin{aligned} R(\eta|\alpha,\mu) &= \mathbf{Y}^T(M - M_0)\mathbf{Y} \\ R(\eta|\mu) &= \mathbf{Y}^T(M_1 - M_0)\mathbf{Y} \end{aligned}$$

# Confidence Regions

$100(1-\alpha)\%$ Confidence Region(CR) for $\Lambda^T\beta$ consists of all the vectors $d$ satisfying the inequality

$$\frac{[\Lambda^T\hat{\beta}-d]^T[\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda]^-[\Lambda^T\hat{\beta}-d]/r(\Lambda)}{MSE} \leq (1-\alpha, r(\Lambda), r(\mathbf{I}-M))$$

These vectors form an ellipsoid in $r(\Lambda)$-dimensional space.

For regression problems, if we take $P^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, then $\Lambda^T\beta = P^T\mathbf{X}\beta = \beta = d$. The $100(1-\alpha)\%$ CR for $\beta$ is

$$\begin{aligned} & \frac{[\Lambda^T\hat{\beta}-d]^T[\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda]^-[\Lambda^T\hat{\beta}-d]/r(\Lambda)}{MSE} \\ = \quad & \frac{(\hat{\beta}-\beta)^T(\mathbf{X}^T\mathbf{X})(\hat{\beta}-\beta)/p}{MSE} \leq (1-\alpha, p, n-p) \end{aligned}$$

# Tests for Generalized Least Squares Models

Let $V$ be a known positive definite and also let $V = QQ^T$. With $\mathcal{C}(X_0) \subset \mathcal{C}(\mathbf{X})$,

$$
\begin{aligned}
\mathbf{Y} &= \mathbf{X}\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 V) && (1)\\
Q^{-1}\mathbf{Y} &= Q^{-1}\mathbf{X}\beta + Q^{-1}\epsilon, \quad Q^{-1}\epsilon \sim N(0, \sigma^2 \mathbf{I}) && (2)\\
\mathbf{Y} &= \mathbf{X}_0\beta_0 + \epsilon, \quad \epsilon \sim N(0, \sigma^2 V) && (3)\\
Q^{-1}\mathbf{Y} &= Q^{-1}\mathbf{X}_0\beta_0 + Q^{-1}\epsilon, \quad Q^{-1}\epsilon \sim N(0, \sigma^2 \mathbf{I}) && (4)
\end{aligned}
$$

$$\text{Testing (3) vs (1)} \Longleftrightarrow \text{Testing (4) vs (2)}$$

**NOTE:**

$$\mathcal{C}(Q^{-1}X_0) \subset \mathcal{C}(Q^{-1}\mathbf{X})$$

# Tests for Generalized Least Squares Models

From Section 2.7,

$$
\begin{aligned}
A &= \mathbf{X}(\mathbf{X}^T V^{-1}\mathbf{X})^{-}\mathbf{X}^T V^{-1} \\
\text{MSE} &= \mathbf{Y}^T(\mathbf{I}-A)^T V^{-1}(\mathbf{I}-A)\mathbf{Y}/(n-r(\mathbf{X})) \\
A_0 &= \mathbf{X}_0(\mathbf{X}_0^T V^{-1}\mathbf{X}_0)^{-}\mathbf{X}_0^T V^{-1}
\end{aligned}
$$

**Theorem 3.8.1**

(i) $$\frac{\frac{\mathbf{Y}^T(A-A_0)V^{-1}(A-A_0)\mathbf{Y}}{r(\mathbf{X})-r(\mathbf{X}_0)}}{MSE} \sim F(r(\mathbf{X})-r(\mathbf{X}_0), n-r(\mathbf{X}), \delta^2)$$

where $\delta^2 = \beta^T\mathbf{X}^T(A-A_0)V^{-1}(A-A_0)\mathbf{X}\beta/2\sigma^2$

(ii) $$\beta^T\mathbf{X}^T(A-A_0)V^{-1}(A-A_0)\mathbf{X}\beta = 0 \iff \mathrm{E}(\mathbf{Y}) \in \mathcal{C}(\mathbf{X}_0)$$

# Tests for Generalized Least Squares Models

**Theorem 3.8.2** Let $\Lambda^T\beta$ be estimable. Then the test statistic for $H_0 : \Lambda^T\beta = 0$ is

$$\text{(i)} \quad \frac{\hat{\beta}^T\Lambda[\Lambda^T(\mathbf{X}^T V^{-1}\mathbf{X})^-\Lambda]^-\Lambda^T\hat{\beta}/r(\Lambda)}{MSE} \sim F(r(\lambda), n - r(\mathbf{X}), \delta^2)$$

where $\delta^2 = \beta^T\Lambda[\Lambda^T(\mathbf{X}^T V^{-1}\mathbf{X})^-\Lambda]^-\Lambda^T\beta/2\sigma^2$

$$\text{(ii)} \quad \beta^T\Lambda[\Lambda^T(\mathbf{X}^T V^{-1}\mathbf{X})^-\Lambda]^-\Lambda^T\beta = 0 \Longleftrightarrow \Lambda^T\beta = 0$$

# Tests for Generalized Least Squares Models

**Theorem 3.8.3**

(i) $$\frac{\mathbf{Y}^T(A - A_0)V^{-1}(A - A_0)\mathbf{Y}}{\sigma^2} \sim \chi^2(r(\mathbf{X}) - r(\mathbf{X}_0), \delta^2)$$

where $\delta^2 = \beta^T\mathbf{X}^T(A - A_0)V^{-1}(A - A_0)\mathbf{X}\beta/2\sigma^2$ and

$$\delta^2 = 0 \iff \mathrm{E}(\mathbf{Y}) \in \mathcal{C}(\mathbf{X}_0)$$

(ii) $$\hat{\beta}^T\Lambda[\Lambda^T(\mathbf{X}^TV^{-1}\mathbf{X})^-\Lambda]^-\Lambda^T\hat{\beta}/2\sigma^2\chi^2(r(\Lambda), \delta^2)$$

where $\delta^2 = \beta^T\Lambda[\Lambda^T(\mathbf{X}^TV^{-1}\mathbf{X})^-\Lambda]^-\Lambda^T\beta/2\sigma^2$ and

$$\delta^2 = 0 \iff \Lambda^T\beta = 0$$