STAT 574 Linear and Nonlinear Mixed Models

Lecture 3: Statistical Properties of Linear Mixed Models

Chencheng Cai

Washington State University

Identifiability

- Statistical model: a family of distributions for $\boldsymbol{y}$ parametrized by the vector $\boldsymbol{\theta}$ $\{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$.
- The model is **identifiable** if

$$P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2} \longrightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

- or, equivalently, if

$$\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \longrightarrow P_{\boldsymbol{\theta}_1} \neq P_{\boldsymbol{\theta}_2}$$

Identifiability — Example

- Consider a linear regression model:

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

- Is the above model Identifiable?

Identifiability — Example

- Consider a linear regression model:

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

- Is the above model Identifiable?
- (counter-)example for Identifiability:

$$\begin{aligned}
& p(\boldsymbol{y} \mid \boldsymbol{\beta}_1, \sigma_1^2) = p(\boldsymbol{y} \mid \boldsymbol{\beta}_2, \sigma_2^2) \longrightarrow \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 \text{ and } \sigma_1^2 = \sigma_2^2 \\
\Longrightarrow & \boldsymbol{X}\boldsymbol{\beta}_1 = \boldsymbol{X}\boldsymbol{\beta}_2 \text{ and } \sigma_1^2 = \sigma_2^2 \longrightarrow \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 \text{ and } \sigma_1^2 = \sigma_2^2 \\
\Longrightarrow & \boldsymbol{X}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) = \mathbf{0} \longrightarrow \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2
\end{aligned}$$

  The last implication requires that $\boldsymbol{X}$ is of full (column) rank.
- The linear regression model is identifiable when $\boldsymbol{X}$ is of full rank.

Identifiability — Example

- Consider the following model of $\boldsymbol{Y}$: (low-rank approximation of matrix)

$$\boldsymbol{Y} = \boldsymbol{u}\boldsymbol{v}^T + \sigma^2 \boldsymbol{E},$$

  where $\boldsymbol{u}\boldsymbol{v}^T$ is an unknown rank-one signal part and $\boldsymbol{E}$ is a noise matrix with IID standard Gaussian entries.
- Is the above model identifiable?

Identifiability — Example

- Consider the following model of $\boldsymbol{Y}$: (low-rank approximation of matrix)

  $$\boldsymbol{Y} = \boldsymbol{u}\boldsymbol{v}^T + \sigma^2 \boldsymbol{E},$$

  where $\boldsymbol{u}\boldsymbol{v}^T$ is an unknown rank-one signal part and $\boldsymbol{E}$ is a noise matrix with IID standard Gaussian entries.
- Is the above model identifiable?
- (counter-)examples:
  - $\boldsymbol{u} \to -\boldsymbol{u}$ and $\boldsymbol{v} \to -\boldsymbol{v}$.
  - $\boldsymbol{u} \to c\boldsymbol{u}$ and $\boldsymbol{v} \to c^{-1}\boldsymbol{v}$ for any $c \neq 0$.
- **Identifiability Conditions** (conditions that make the model identifiable)
  - (reparametrization) $\boldsymbol{Y} = \lambda \boldsymbol{u}\boldsymbol{v}^T + \sigma^2 \boldsymbol{E}$
  - (normalization) $\|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1$ and $\lambda > 0$.
  - (sign fixing) The first non-zero element in $\boldsymbol{u}$ is positive.

Identifiability — sufficient and necessary conditions

For regression models with **normal** distribution, a sufficient and necessary condition for Identifiability is

$$\mathbb{E}_{\boldsymbol{\theta}_1}(\boldsymbol{y}) = \mathbb{E}_{\boldsymbol{\theta}_2}(\boldsymbol{y}) \text{ and } \mathrm{Cov}_{\boldsymbol{\theta}_1}(\boldsymbol{y}) = \mathrm{Cov}_{\boldsymbol{\theta}_2}(\boldsymbol{y}) \longrightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

- Let a regression model be defined as

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{f}(\boldsymbol{\beta}), \boldsymbol{V}(\boldsymbol{\beta}, \boldsymbol{\theta})).$$

- Identifiability condition:

$$\boldsymbol{f}(\boldsymbol{\beta}_1) = \boldsymbol{f}(\boldsymbol{\beta}_2) \text{ and } \boldsymbol{V}(\boldsymbol{\beta}_1, \boldsymbol{\theta}_1) = \boldsymbol{V}(\boldsymbol{\beta}_2, \boldsymbol{\theta}_2) \longrightarrow \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2, \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2.$$

Identifiability Conditions for LME

Theorem

If matrix $\boldsymbol{X}$ has full rank, at least one matrix $\boldsymbol{Z}_i$ has full rank, and $\sum_{i=1}^{N}(n_i - k) > 0$, the LME model is identifiable.

- $\boldsymbol{X} = [\boldsymbol{X}_1^T, \boldsymbol{X}_2^T, \cdots, \boldsymbol{X}_N^T]^T$
- k is the number of random effects.
- n_i is the sample size for group i.

Identifiability Conditions for LME — Proof

Recall the model:

$$\boldsymbol{y} = \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \text{diag}(\boldsymbol{V}_1, \cdots, \boldsymbol{V}_N)),$$

where $\boldsymbol{V}_i = \boldsymbol{I}_{n_i} + \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T$.

(1) Identifiability for β

When $\boldsymbol{X}$ has full rank, $\boldsymbol{X}\boldsymbol{\beta}_1 = \boldsymbol{X}\boldsymbol{\beta}_2$ implies $\boldsymbol{X}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) = \mathbf{0}$, which gives $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

(2) Identifiability for D and σ^2

Need to show that when at least one matrix $\boldsymbol{Z}_i$ has full rank, and $\sum(n_i - k) > 0$, we have

$$\sigma_1^2(\boldsymbol{I}_{n_i} + \boldsymbol{Z}_i \boldsymbol{D}_1 \boldsymbol{Z}_i^T) = \sigma_2^2(\boldsymbol{I}_{n_i} + \boldsymbol{Z}_i \boldsymbol{D}_2 \boldsymbol{Z}_i^T) \; \forall i = 1, \ldots, N \;\longrightarrow\; \sigma_1^2 = \sigma_2^2, \; \boldsymbol{D}_1 = \boldsymbol{D}_2.$$

Identifiability Conditions for LME — Proof

(2) Identifiability for D and σ^2

Let $\delta = \sigma_1^2 - \sigma_2^2$ and $\boldsymbol{\Delta} = \sigma_1^2 \boldsymbol{D}_1 - \sigma_2^2 \boldsymbol{D}_2$. So we have

$$\delta \boldsymbol{I}_{n_i} + \boldsymbol{Z}_i \boldsymbol{\Delta} \boldsymbol{Z}_i^T = \mathbf{0} \ \forall i. \tag{1}$$

Without loss of generality, assume $\boldsymbol{Z}_1$ has full rank. Then

$$\delta I_{n_1} + \boldsymbol{Z}_1 \boldsymbol{\Delta} \boldsymbol{Z}_1^T = \mathbf{0} \quad \Longrightarrow \quad \delta(\boldsymbol{Z}_1^T \boldsymbol{Z})^{-1} + \boldsymbol{\Delta} = \mathbf{0} \tag{2}$$

Therefore, by substitute $\boldsymbol{\Delta}$ in (1), we have

$$\delta \boldsymbol{I}_{n_i} - \delta \boldsymbol{Z}_i (\boldsymbol{Z}_1^T \boldsymbol{Z}_1)^{-1} \boldsymbol{Z}_i^T = \mathbf{0} \ \forall i \tag{3}$$

Identifiability Conditions for LME — Proof

Because $\sum(n_i - k) > 0$, there exists l such that $n_l - k > 0$. Then from (3), we know

$$\delta \boldsymbol{I}_{n_l} - \delta \boldsymbol{Z}_l (\boldsymbol{Z}_1^T \boldsymbol{Z}_1)^{-1} \boldsymbol{Z}_l^T = \mathbf{0}$$

Notice that

$$\mathrm{rank}(\boldsymbol{Z}_l (\boldsymbol{Z}_1^T \boldsymbol{Z}_1)^{-1} \boldsymbol{Z}_l^T) \leq \mathrm{rank}(\boldsymbol{Z}_l) \leq k < n_l.$$

Hence, $\boldsymbol{I}_{n_l} - \boldsymbol{Z}_l (\boldsymbol{Z}_1^T \boldsymbol{Z}_1)^{-1} \boldsymbol{Z}_l^T \neq \mathbf{0}$. Then we must have $\delta = 0$. Furthermore, from (2), we know $\boldsymbol{\Delta} = \mathbf{0}$. $\delta = 0$ and $\boldsymbol{\Delta} = \mathbf{0}$ imply $\sigma_1^2 = \sigma_2^2$ and $\boldsymbol{D}_1 = \boldsymbol{D}_2$.

Remark: the textbook proves the theorem using the positive definiteness of the information matrix, which overkills the problem.

Identifiability is not the validity of a model

Consider a random-coefficient model:

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

where $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2\boldsymbol{D})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.
We can write

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

where $\boldsymbol{\delta} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{D})$.
Or, equivalently,

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2(\boldsymbol{I} + \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^T)).$$

Is this a mixed-effect model? Is this a balanced random coefficient model?

Identifiability is not the validity of a model

The complete profiled log-likelihood function of $\boldsymbol{D}$ is (verify!)

$$\ell_p(\boldsymbol{D}) = -\frac{1}{2}\log|\boldsymbol{I} + \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^T| + C$$

This leads to the trivial solution

$$\hat{\boldsymbol{D}} = \mathbf{0}.$$

- The data is not large enough to identify the "randomness" of $\boldsymbol{b}$.
- We cannot seperate the noise $\boldsymbol{\delta}$ from the mean $\boldsymbol{\beta}$ when having only one group.
- The solution is not trivial if we have more than 1 group.

Information Matrix

Log-likelihood for group i:

$$\ell_i = -\frac{1}{2}\left\{n \log \sigma^2 + \sigma^{-2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{V}_i^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \log|\boldsymbol{V}_i|\right\}$$

Derivatives:

$$\frac{\partial \ell_i}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sigma^{-2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{V}_i^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^2} \sim -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2}\chi_n^2$$

and

$$\mathrm{Var}\left(\frac{\partial \ell_1}{\partial \sigma^2}\right) = \frac{n}{2\sigma^4}$$

Information Matrix

Log-likelihood for group i:

$$\ell_i = -\frac{1}{2}\left\{n \log \sigma^2 + \sigma^{-2}(\boldsymbol{y} - \boldsymbol{X\beta})^T \boldsymbol{V}_i^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) + \log|\boldsymbol{V}_i|\right\}$$

Derivatives:

$$\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \sigma^{-2}\boldsymbol{X}^T\boldsymbol{V}_i^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})$$

and $\frac{\partial \ell_i}{\partial \boldsymbol{\beta}}$ is uncorrelated with $\frac{\partial \ell_i}{\partial \sigma^2}$ and $\frac{\partial \ell_i}{\partial \boldsymbol{D}}$ (check textbook). Therefore,

$$\operatorname{cov}\left(\hat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2 \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T(\boldsymbol{I} + \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i^T)^{-1}\boldsymbol{X}_i\right)^{-1}$$

- $\sum_{i=1}^{N}$ because we consider N clusters.
- $()^{-1}$ because the information matrix is block diagonal.
- It is the asymptotic variance — we used the information matrix.

Information Matrix

Some useful facts (under certain regulation conditions):

- $\mathbb{E}[\partial \ell / \partial \boldsymbol{\theta}] = \mathbf{0}$.
- $\boldsymbol{I}(\theta) = \mathbb{E}[\partial^2 \ell / \partial \boldsymbol{\theta}^2]$.
- $[\boldsymbol{I}(\theta)]_{ij} = \mathbb{E}[\partial^2 \ell / (\partial \theta_i \partial \theta_j)] = \mathbb{E}[(\partial \ell / \partial \theta_i)(\partial \ell / \partial \theta_j)] = \mathrm{Cov}[(\partial \ell / \partial \theta_i), (\partial \ell / \partial \theta_j)]$
- $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \boldsymbol{I}^{-1}$

Wald Confidence Interval

Let s_j be the variance of $\hat{\beta}_j$, that is $[\mathrm{cov}(\hat{\boldsymbol{\beta}})]_{jj}$. Asymptotically, we have

$$\frac{\hat{\beta}_j}{s_j} \sim \mathcal{N}(0,1).$$

Or finitely, we have

$$\frac{\hat{\beta}_j}{s_j} \sim t_{N_T - m},$$

where m is the number of fixed effects.

Warning: $\hat{\beta}_j / s_j$ is NOT t-distributed. **Why?**

Wald Confidence Interval

Let s_j be the variance of $\hat{\beta}_j$, that is $[\mathrm{cov}(\hat{\boldsymbol{\beta}})]_{jj}$. Asymptotically, we have

$$\frac{\hat{\beta}_j}{s_j} \sim \mathcal{N}(0,1).$$

Or finitely, we have

$$\frac{\hat{\beta}_j}{s_j} \sim t_{N_T - m},$$

where m is the number of fixed effects.

Warning: $\hat{\beta}_j / s_j$ is NOT t-distributed. **Why?** $\boldsymbol{D}$ is estiamted.

Profile-likelihood Confidence Interval

Assume the parameter has two parts $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$.
Consider the hypothesis testing:

$$H_0 : \boldsymbol{\theta}_0 = \boldsymbol{\theta}^* \quad H_a : \boldsymbol{\theta}_0 \neq \mathbf{0}$$

Likelihood ratio test:

$$LR = \frac{\max_{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1} \ L(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)}{\max_{\boldsymbol{\theta}_1} \ L(\boldsymbol{\theta}^*, \boldsymbol{\theta}_1)}$$

Distribution:

$$2 \log LR \sim \chi^2_{\dim(\boldsymbol{\theta}_0)}$$

Reject when:

$$\max_{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) - \max_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}^*, \boldsymbol{\theta}_1) > \frac{1}{2} \chi^2_{1-\alpha, \dim(\boldsymbol{\theta}_0)}$$

For what values of θ^*, we cannot rejct?

Profile-likelihood Confidence Interval

Assume the parameter has two parts $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$.
Consider the hypothesis testing:

$$H_0 : \boldsymbol{\theta}_0 = \boldsymbol{\theta}^* \quad H_a : \boldsymbol{\theta}_0 \neq \mathbf{0}$$

Likelihood ratio test:

$$LR = \frac{\max_{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1} \ L(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)}{\max_{\boldsymbol{\theta}_1} \ L(\boldsymbol{\theta}^*, \boldsymbol{\theta}_1)}$$

Distribution:

$$2 \log LR \sim \chi^2_{\dim(\boldsymbol{\theta}_0)}$$

Reject when:

$$\max_{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) - \max_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}^*, \boldsymbol{\theta}_1) > \frac{1}{2} \chi^2_{1-\alpha, \dim(\boldsymbol{\theta}_0)}$$

For what values of θ^*, we cannot rejct?

$$\{\boldsymbol{\theta}^* : \max_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}^*, \boldsymbol{\theta}_1) \geq \max_{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) - \frac{1}{2} \chi^2_{1-\alpha, \dim(\boldsymbol{\theta}_0)}\}$$

Profile-likelihood Confidence Interval

Profile-likelihood Confidence Interval for θ_0 (one of the parameters):

$$\{\theta : \max_{\boldsymbol{\theta}_1} \ell(\theta, \boldsymbol{\theta}_1) \geq \max_{\theta_0, \boldsymbol{\theta}_1} \ell(\theta_0, \boldsymbol{\theta}_1) - \frac{1}{2}\chi^2_{1-\alpha,1}\}$$

Or $\theta_L < \theta_R$ are solutions to

$$\max_{\boldsymbol{\theta}_1} \ell(x, \boldsymbol{\theta}_1) = \max_{\theta_0, \boldsymbol{\theta}_1} \ell(\theta_0, \boldsymbol{\theta}_1) - \frac{1}{2} Z^2_{1-\alpha/2},$$

where $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of a standard normal distribution. The PL confidence interval is (θ_L, θ_R). **Version adjusted for d.f.** $\theta_L < \theta_R$ are solutions to

$$\max_{\boldsymbol{\theta}_1} \ell(x, \boldsymbol{\theta}_1) = \max_{\theta_0, \boldsymbol{\theta}_1} \ell(\theta_0, \boldsymbol{\theta}_1) - \frac{1}{2} t^2_{1-\alpha/2, n-m},$$

where $t^2_{1-\alpha/2,n-m}$ is the $(1-\alpha/2)$-quantile of a t-distribution with d.f. $(n-m)$.

Profile-likelihood Confidence Interval for LME

- Let $\boldsymbol{\beta}_{-j}$ be the fixed effect coefficient vector except the j-th one.
- PL CI:

$$\max_{\boldsymbol{\beta}_{-j},\boldsymbol{D},\sigma^2} \ell(\beta_j, \boldsymbol{\beta}_{-j}, \boldsymbol{D}, \sigma^2) = \max_{\boldsymbol{\beta},\boldsymbol{D},\sigma^2} \ell(\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) - \frac{1}{2} Z^2_{1-\alpha/2}$$

- How to calculate $\max_{\boldsymbol{\beta},\boldsymbol{D},\sigma^2} \ell(\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)$?

Profile-likelihood Confidence Interval for LME

- Let $\boldsymbol{\beta}_{-j}$ be the fixed effect coefficient vector except the j-th one.
- PL CI:

$$\max_{\boldsymbol{\beta}_{-j},\boldsymbol{D},\sigma^2} \ell(\beta_j, \boldsymbol{\beta}_{-j}, \boldsymbol{D}, \sigma^2) = \max_{\boldsymbol{\beta},\boldsymbol{D},\sigma^2} \ell(\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) - \frac{1}{2} Z^2_{1-\alpha/2}$$

- How to calculate $\max_{\boldsymbol{\beta},\boldsymbol{D},\sigma^2} \ell(\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)$? MLE.
- How to get $\max_{\boldsymbol{\beta}_{-j},\boldsymbol{D},\sigma^2} \ell(\beta_j, \boldsymbol{\beta}_{-j}, \boldsymbol{D}, \sigma^2)$?

Profile-likelihood Confidence Interval for LME

- Let $\boldsymbol{\beta}_{-j}$ be the fixed effect coefficient vector except the j-th one.
- PL CI:

$$\max_{\boldsymbol{\beta}_{-j}, \boldsymbol{D}, \sigma^2} \ell(\beta_j, \boldsymbol{\beta}_{-j}, \boldsymbol{D}, \sigma^2) = \max_{\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2} \ell(\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2) - \frac{1}{2} Z^2_{1-\alpha/2}$$

- How to calculate $\max_{\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2} \ell(\boldsymbol{\beta}, \boldsymbol{D}, \sigma^2)$? MLE.
- How to get $\max_{\boldsymbol{\beta}_{-j}, \boldsymbol{D}, \sigma^2} \ell(\beta_j, \boldsymbol{\beta}_{-j}, \boldsymbol{D}, \sigma^2)$? MLE for the following LME model:

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{X}}_i \boldsymbol{\beta}_{-j} + \boldsymbol{Z}_i \boldsymbol{b} + \epsilon_i$$

  - $\tilde{\boldsymbol{y}}_i = \boldsymbol{y}_i - \beta_j \boldsymbol{x}^j$
  - $\tilde{\boldsymbol{X}}$: removing j-th column from $\boldsymbol{X}_i$.

F-test for $\boldsymbol{D}$

Consider the test:

$$\boldsymbol{D} = \mathbf{0}$$

Possible test?

- likelihood ratio test? No. The limit distribution of LR is not χ^2 because $\boldsymbol{D} = \mathbf{0}$ is on the boundary of the parameter space.
- sum of squares? Yes. F-test.

F-test for D

- Without random effect terms:

$$S_{OLS} = \sum_{i=1}^{N} \left\| \boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}_{OLS} \right\|^2 = \min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2$$

- With random effect terms:

$$S_{min} = \min_{\boldsymbol{\gamma}} \|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\gamma}\|^2,$$

where $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_N^T)^T$, $\boldsymbol{\gamma} = (\boldsymbol{\beta}^T, \boldsymbol{b}_1^T, \ldots, \boldsymbol{b}_N^T)^T$, and

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{Z}_1 & 0 & \cdots & 0 \\ \boldsymbol{X}_2 & 0 & \boldsymbol{Z}_2 & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ \boldsymbol{X}_N & 0 & 0 & \vdots & \boldsymbol{Z}_N \end{bmatrix}$$

F-test for $\boldsymbol{D}$

Theorem

Let $r = \mathrm{rank}(\boldsymbol{W})$. *When* $\boldsymbol{D} = \mathbf{0}$, *we have*

$$\frac{(S_{OLS} - S_{min})/(r - m)}{S_{min}/(N_T - r)} \sim F(r - m, N_T - r).$$

Proof: Because $\boldsymbol{X}$ is a submatrix of $\boldsymbol{W}$. The test is the same as testing

$$H_0 : \boldsymbol{b}_1 = \cdots = \boldsymbol{b}_N = \mathbf{0}$$

F-test for D — Special Case

VARCOMP model:

$$y_{ij} = \beta + b_i + \epsilon_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, N,$$

where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and $b_i \sim \mathcal{N}(0, \sigma^2 d)$.
The F-test:

$$\frac{\sum n_i(\bar{y}_i - \sum_j n_j \bar{y}_j / N_T)^2 / (N-1)}{\left(\sum_{ij} y_{ij}^2 - \sum_i n_i \bar{y}_i^2\right) / (N_T - N)} \sim F(N-1, N_T - N)$$

- numerator: between-group variation
- denominator: within-group variation

Finite Sample Properties for MLE

Is $\hat{\beta}$ unbiased?

Recall:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{y}_i \right),$$

where $\hat{\boldsymbol{V}}_i = \boldsymbol{I} + \boldsymbol{Z}_i \hat{\boldsymbol{D}} \boldsymbol{Z}_i^T$ is the estimated covariance matrix for group i.

Finite Sample Properties for MLE

Is $\hat{\beta}$ unbiased?

Recall:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1} \boldsymbol{y}_i\right),$$

where $\hat{\boldsymbol{V}}_i = \boldsymbol{I} + \boldsymbol{Z}_i \hat{\boldsymbol{D}} \boldsymbol{Z}_i^T$ is the estimated covariance matrix for group i.

- all randomness come from $\boldsymbol{\eta}_i$ for $i = 1, \ldots, N$. $(\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\eta}_i)$
- $\boldsymbol{\eta}_i$ is symmetric around $\mathbf{0}$: $p(\boldsymbol{\eta}_i) = p(-\boldsymbol{\eta}_i)$.
- $\hat{\boldsymbol{D}}$ is an even function of $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_N$: $\hat{\boldsymbol{D}}(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_N) = \hat{\boldsymbol{D}}(-\boldsymbol{\eta}_1, \ldots, -\boldsymbol{\eta}_N)$. Why? The profiled LLH of $\boldsymbol{D}$ is an even function of $\boldsymbol{D}$.
- Therefore, the distribution of $\hat{\boldsymbol{\beta}}$ is symmetric at $\boldsymbol{\beta}$.

Finite Sample Properties for MLE

A quick proof: (let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \ldots, \boldsymbol{\eta}_N^T)^T$)

$$\begin{aligned}\mathbb{E}[\hat{\boldsymbol{\beta}}] &= \int \hat{\boldsymbol{\beta}}(\boldsymbol{\eta}) p(\boldsymbol{\eta}) d\boldsymbol{\eta} = \int \hat{\boldsymbol{\beta}}(\boldsymbol{\eta}) \frac{p(\boldsymbol{\eta}) + p(-\boldsymbol{\eta})}{2} d\boldsymbol{\eta} \\ &= \frac{1}{2} \int \hat{\boldsymbol{\beta}}(\boldsymbol{\eta}) p(\boldsymbol{\eta}) d\boldsymbol{\eta} - \frac{1}{2} \int \hat{\boldsymbol{\beta}}(-\boldsymbol{\eta}) p(\boldsymbol{\eta}) d\boldsymbol{\eta} = \int \frac{\hat{\boldsymbol{\beta}}(\boldsymbol{\eta}) + \hat{\boldsymbol{\beta}}(-\boldsymbol{\eta})}{2} p(\boldsymbol{\eta}) d\boldsymbol{\eta}.\end{aligned}$$

Because we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\boldsymbol{\eta}) + \hat{\boldsymbol{\beta}}(-\boldsymbol{\eta}) = &\left(\sum_{i=1}^N \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{\eta}) \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^N \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{\eta}) (\boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\eta}) \right) \\ &- \left(\sum_{i=1}^N \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1}(-\boldsymbol{\eta}) \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^N \boldsymbol{X}_i^T \hat{\boldsymbol{V}}_i^{-1}(-\boldsymbol{\eta}) (\boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{\eta}) \right) = 2\boldsymbol{\beta},\end{aligned}$$

$\hat{\boldsymbol{\beta}}$ is unbiased.

Finite Sample Properties for MLE

Are the following variance estimators unbiased?

- $\hat{\sigma}^2_{ML}$
- $\hat{\boldsymbol{D}}_{ML}$
- $\hat{\sigma}^2_{ML}\hat{\boldsymbol{D}}_{ML}$
- $\hat{\sigma}^2_{REML}$
- $\hat{\boldsymbol{D}}_{REML}$
- $\hat{\sigma}^2_{REML}\hat{\boldsymbol{D}}_{REML}$

Finite Sample Properties for MLE

Are the following variance estimators unbiased?

- $\hat{\sigma}^2_{ML}$ **biased**
- $\hat{\boldsymbol{D}}_{ML}$ **biased**
- $\hat{\sigma}^2_{ML}\hat{\boldsymbol{D}}_{ML}$ **biased**
- $\hat{\sigma}^2_{REML}$ **unbiased**
- $\hat{\boldsymbol{D}}_{REML}$ **unbiased**
- $\hat{\sigma}^2_{REML}\hat{\boldsymbol{D}}_{REML}$ **N/A**

Finite Sample Properties for MLE

A special case:

- $\hat{\sigma}^2_{ML}$ and $\hat{\sigma}^2_{ML}\hat{\boldsymbol{D}}_{REML}$ are unbiased for the balanced random-coefficient models.
- Recall:

$$\hat{\sigma}^2_{ML} = \hat{\sigma}^2_{RML} = \frac{1}{N(n-m)} \sum_{i=1}^{N} \boldsymbol{y}_i^T (\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T)\boldsymbol{y}_i$$

$$\hat{\boldsymbol{D}}_{RML} = \frac{1}{(N-1)\hat{\sigma}^2_{ML}} (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\hat{\boldsymbol{E}}\hat{\boldsymbol{E}}^T\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} - (\boldsymbol{Z}^T\boldsymbol{Z})^{-1}$$

Finite Sample Properties for MLE

- $\hat{\sigma}^2_{ML}$ and $\hat{\sigma}^2_{ML}\hat{\boldsymbol{D}}_{REML}$ are unbiased for the balanced random-coefficient models.

Proof:

$$\mathbb{E}[\hat{\sigma}^2_{ML}] = \frac{1}{N(n-m)} \sum_{i=1}^{N} \mathbb{E}[\boldsymbol{e}_i^T(\boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T)\boldsymbol{e}_i] = \sigma^2$$

$$\begin{aligned}\mathbb{E}[\hat{\sigma}^2_{ML}\hat{\boldsymbol{D}}_{REML}] &= \frac{1}{N-1}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\mathbb{E}[\hat{\boldsymbol{E}}\hat{\boldsymbol{E}}^T]\boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} - \sigma^2(\boldsymbol{Z}^T\boldsymbol{Z})^{-1} \\ &= \sigma^2\boldsymbol{D}\end{aligned}$$

Large Sample Properties — Deterministic v.s. Stochastic Schemes

Example: (linear regression model)

$$y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \epsilon_i.$$

and

$$\hat{\boldsymbol{\beta}}_n = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i^T y_i \right)$$

Large Sample Properties — Deterministic v.s. Stochastic Schemes

$$\hat{\boldsymbol{\beta}}_n = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i^T y_i\right)$$

Deterministic Scheme:
If

$$\sup \|\boldsymbol{x}_i\| \leq B \text{ and } \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T = \boldsymbol{A}$$

then,

$$\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 \boldsymbol{A}^{-1})$$

Large Sample Properties — Deterministic v.s. Stochastic Schemes

$$\hat{\boldsymbol{\beta}}_n = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i^T y_i \right)$$

Stochastic Scheme:

If $\boldsymbol{x}_i$'s are i.i.d. with mean μ_x and covariance $\boldsymbol{V}_x$, then,

$$\hat{\boldsymbol{\beta}}_n \xrightarrow{P} \boldsymbol{\beta}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 \boldsymbol{V}_x^{-1})$$

Large Sample Properties — LME in Stochastic Schemes

We assume groups $(\boldsymbol{X}_i, \boldsymbol{Z}_i, n_i) \sim f(\cdot \mid \xi)$ are i.i.d distributed. The full model:

$$\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2(\boldsymbol{I} + \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^T)), \quad (\boldsymbol{X}_i, \boldsymbol{Z}_i, n_i) \sim f(\cdot \mid \xi)$$

The full log-likelihood:

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) + \sum_{i=1}^{N} \log f(\boldsymbol{X}_i, \boldsymbol{Z}_i, n_i \mid \xi)$$

Then we have

$$\sqrt{N}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \sigma^2 \left(\sum_{j=1}^{\infty} p_j \mathbb{E}[\boldsymbol{Z}_1\boldsymbol{Z}_1^T \mid n_1 = j]\right)^{-1} + \sigma^2\boldsymbol{D}\right)$$

Large Sample Properties — LME in Stochastic Schemes

What if $\inf_i n_i \to \infty$ but N constant?

Large Sample Properties — LME in Stochastic Schemes

What if $\inf_i n_i \to \infty$ but N constant?

- $n_i \to \infty$: we have all possible observations for group i (with $\boldsymbol{b}_i$)
- Same to the previous argument: we have near-noiseless observations for group i
- A finite number of $\boldsymbol{b}_i$'s do not guarantee consistency in estimating $\boldsymbol{D}$.
- Condition: Not consistent for $\boldsymbol{D}$.

Large Sample Properties — Equivalence of ML and REML

- The log-likelihood functions for ML and REML differ by

$$\log \left| \sum_{i=1}^{N} \boldsymbol{X}_i^T (\boldsymbol{I} + \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T)^{-1} \boldsymbol{X}_i \right|$$

- The term is $o_p(N)$ whereas other terms in log-likelihood functions are $O_p(N)$.
- The difference in log-likelihood functions vanishes when $N \to \infty$.

Large Sample Properties — Equivalence of ML and REML

- Need to show

$$\lim_{N\to\infty} \frac{1}{N} \frac{\partial}{\partial \boldsymbol{D}} \log \left| \sum_{i=1}^{N} \boldsymbol{X}_i^T (\boldsymbol{I} + \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T)^{-1} \boldsymbol{X}_i \right| = 0$$

Estimation of Random Effects

Now assume we already have the values for $\boldsymbol{\beta}, \sigma^2$ and $\boldsymbol{D}$ (or estimates of them)

How to estimate the random effect coefficients for each group?
i.e. get $\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_N$ given $\boldsymbol{\beta}$, σ^2, $\boldsymbol{D}$.

Estimation of Random Effects — Bayesian Approach

From Bayesian perspective,

- Prior: $\boldsymbol{b}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{D})$.
- Likelihood: $(\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \mid \boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{Z}_i \boldsymbol{b}_i, \sigma^2 \boldsymbol{I})$.
- Posterior: (let $\boldsymbol{\eta} = \boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}$)

$$
\begin{aligned}
p(\boldsymbol{b}_i \mid \boldsymbol{\eta}) &\propto \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{\eta} - \boldsymbol{Z}_i \boldsymbol{b}_i\|^2\right\} \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{b}_i^T (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1}) \boldsymbol{b}_i - 2\boldsymbol{\eta}^T \boldsymbol{Z}_i \boldsymbol{b}_i\right]\right\} \\
&\sim \mathcal{N}((\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{\eta}, \sigma^2 (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1})
\end{aligned}
$$

- The estimate is

$$
\hat{\boldsymbol{b}}_i = (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})
$$

Estimation of Random Effects — Simultaneous Estimation

The fixed effect coefficients and the random effect coefficients can be estimated simultaneously through

$$\min_{\boldsymbol{\beta}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N} \sum_{i=1}^{N} \left[\|\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta} - \boldsymbol{Z}_i \boldsymbol{b}_i\|^2 + \boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i \right]$$

- Optimization with respect to $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$ is the same as in Bayesian approach.
- Plug in the solution for $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N$, and we have an optimization problem for $\boldsymbol{\beta}$.

Estimation of Random Effects — Simultaneous Estimation

Let $\boldsymbol{b}_i = (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{\eta}$, we have the objective function is

$$\text{obj.fun.} = \sum_{i=1}^{N} \Big[\| \left[\boldsymbol{I} - \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \right] \boldsymbol{\eta} \|^2 + \boldsymbol{\eta}^T \boldsymbol{Z}_i (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{D}^{-1} (\boldsymbol{Z}_i^T \boldsymbol{Z}_i + \boldsymbol{D}^{-1})^{-1} \boldsymbol{Z}_i^T \boldsymbol{\eta} \Big]$$

Notice that $(\boldsymbol{D}^{-1} + \boldsymbol{Z}_i^T \boldsymbol{Z}_i)^{-1} = \boldsymbol{D} - \boldsymbol{D} \boldsymbol{Z}_i^T (\boldsymbol{I} + \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T)^{-1} \boldsymbol{Z}_i \boldsymbol{D}$.

$$\text{obj.fun.} = \sum_{i=1}^{N} \left[\| \boldsymbol{V}_i^{-1} \boldsymbol{\eta} \|^2 + \boldsymbol{\eta} \boldsymbol{V}_i^{-1} \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{\eta} \right] = \sum_{i=1}^{N} \boldsymbol{\eta}^T \boldsymbol{V}_i^{-1} \boldsymbol{\eta}$$

Same as in the GLS.

Estimation of Random Effects — BLUE

For simplicity, we consider the following model with one group.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}.$$

Consider linear estimators for $\boldsymbol{b}$, that is, $\hat{\boldsymbol{b}} = \boldsymbol{C}\boldsymbol{y}$.

- Expectation: $\mathbb{E}[\boldsymbol{C}\boldsymbol{y}] = \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta}$.
- MSE: $\mathrm{Var}(\boldsymbol{C}\boldsymbol{y} - \boldsymbol{b}) = \boldsymbol{C}\boldsymbol{C}^T + (\boldsymbol{I} - \boldsymbol{C}\boldsymbol{Z})\boldsymbol{D}(\boldsymbol{I} - \boldsymbol{C}\boldsymbol{Z})^T$
- BLUE: for any $\boldsymbol{p}$,

$$\min_{\boldsymbol{C}} \ \boldsymbol{p}^T[\boldsymbol{C}\boldsymbol{C}^T + (\boldsymbol{I} - \boldsymbol{C}\boldsymbol{Z})\boldsymbol{D}(\boldsymbol{I} - \boldsymbol{C}\boldsymbol{Z})^T]\boldsymbol{p} \quad s.t.\ \boldsymbol{C}\boldsymbol{X} = \boldsymbol{0}$$

- The solution is the same as the Bayesian approach.

Hypothesis Testing on Fixed Effects

Consider a generalized linear regression model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{V})$$

Test

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0} \quad v.s. \quad H_a : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{0}$$

Construct RSS and RSS_0 as

$$\mathrm{RSS} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \quad \mathrm{RSS}_0 = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0)^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0),$$

where $\hat{\boldsymbol{\beta}}$ is GLS without constraint and $\hat{\boldsymbol{\beta}}_0$ is GLS under null.

Hypothesis Testing on Fixed Effects

Construct F-test:

$$\frac{(\mathrm{RSS}_0 - \mathrm{RSS})/q}{\mathrm{RSS}/(n-m)} \sim F_{q,n-m},$$

where n is the number of observations, m is the number of covariates, and q is the number of constraints in H_0.
Why?

Hypothesis Testing on Fixed Effects

Construct F-test:

$$\frac{(\mathrm{RSS}_0 - \mathrm{RSS})/q}{\mathrm{RSS}/(n-m)} \sim F_{q,n-m},$$

where n is the number of observations, m is the number of covariates, and q is the number of constraints in H_0.

Why?

- $\mathrm{RSS} \sim \chi^2_{n-m}$
- $\mathrm{RSS}_0 \sim \chi^2_{n-m+q}$
- The space for $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is a subspace of that for $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_0$.

Programming — Test significance of fixed effect

```r
library(nlme)
fit.lme = lme(fixed=Weight~Height+Sex, random=~1|FamilyID, data=data)
summary(fit.lme)
```

```
        Fixed effects:  Weight ~ Height + Sex
                   Value Std.Error DF    t-value p-value
(Intercept) -54.83871  80.15014 51 -0.6841998  0.4969
Height        2.93276   1.22667 51  2.3908287  0.0205
Sex          24.16578   9.82833 51  2.4587871  0.0174
```

Programming — Get the estimated coefficients

```r
fit.lme = lme(fixed=Weight~Height, random=~1|FamilyID, data=data)
coef(fit.lme)
```

	(Intercept)	Height
1	-185.7838	5.345309
2	-209.4829	5.345309
3	-195.4609	5.345309
4	-226.7483	5.345309
5	-216.9965	5.345309
6	-204.6060	5.345309
7	-211.9348	5.345309
8	-215.6025	5.345309
9	-203.5639	5.345309
10	-212.1451	5.345309
11	-208.4306	5.345309
12	-203.9077	5.345309
13	-214.8966	5.345309

Programming — Get the estimated coefficients

```
fit.lme = lmer(Weight~Height +(1|FamilyID), data=data)
coef(fit.lme)$FamilyID
```

	(Intercept)	Height
1	-185.7838	5.345309
2	-209.4829	5.345309
3	-195.4609	5.345309
4	-226.7483	5.345309
5	-216.9965	5.345309
6	-204.6060	5.345309
7	-211.9348	5.345309
8	-215.6025	5.345309
9	-203.5639	5.345309
10	-212.1451	5.345309
11	-208.4306	5.345309
12	-203.9077	5.345309
13	-214.8966	5.345309

Programming — Test random effects

```
Z = as.matrix(bdiag(split(rep(1, dim(data)[1]), data$FamilyID)))

fit0 = lm(data$Weight ~ data$Height)
fit1 = lm(data$Weight ~ 0 + data$Height + Z)
anova(fit0, fit1)
```

```
Analysis of Variance Table

Model 1: data$Weight ~ data$Height
Model 2: data$Weight ~ 0 + data$Height + Z
  Res.Df   RSS Df Sum of Sq    F  Pr(>F)
1     69 55856
2     52 32122 17     23734 2.26 0.01261 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```