3. The General LME Model Formulation and Common Variance – Covariance structures

Specific learning objectives:

1. Write the random intercept & slope model with one covariate in matrix form.
2. Relate the above with the general LME model formulation.
3. Write and explain the Covariance formula for Y.

The General LME Model Formulation

We will build from the simple cases to obtain in each case the general LME formulation below:

$$Y_i = \overbrace{\underbrace{x_i\beta}_{\text{Population mean}} + \underbrace{z_iU_i}_{\text{Subject specific deviation from population mean}}}^{\text{Subject-specific mean}} + \underbrace{\varepsilon_i}_{\text{Random deviation from subject specific mean}};$$

$$j = 1,\ldots,m_i;\ i = 1,\ldots,n.$$

Matrix Representation
Maxillary Distance Data

Linear random effects model:

- ✓ Random intercept

Linear mixed effects model (random + fixed):

- ✓ Random intercept and slope
- ✓ Random intercept and slope by groups

Matrix form, Random intercept model.

$$Y_{ij} = \beta_0 + \beta_1 Age_{ij} + u_{i1} + \varepsilon_{ij}; \quad j = 1,\ldots,4 \quad i = 1,\ldots,27.$$

Or…

$$\begin{aligned}
Y_{i1} &= \beta_0 + \beta_1 Age_{i1} + u_{i1} + \varepsilon_{i1} \\
Y_{i2} &= \beta_0 + \beta_1 Age_{i2} + u_{i1} + \varepsilon_{i2} \\
Y_{i3} &= \beta_0 + \beta_1 Age_{i3} + u_{i1} + \varepsilon_{i3} \\
Y_{i4} &= \beta_0 + \beta_1 Age_{i4} + u_{i1} + \varepsilon_{i4}
\end{aligned}$$

If Age was centered to 11 yrs, the elements (8,10,12,14) would be replaced by (-3,-1,1,3).

Matrix representation: $Y_i = x_i\beta + z_i U_i + \varepsilon_i; \quad i = 1,\ldots,27.$

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix}; \qquad x_i\beta = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix};$$

$U_i = u_{i1}$

$$z_i U_i = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} u_{i1}; \qquad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}.$$

Stacking observations from one individual

Matrix form, Random intercept and slope model.

$$Y_{ij} = \beta_0 + \beta_1 Age_{ij} + u_{i1} + u_{i2} Age_{ij} + \varepsilon_{ij}; \quad j = 1,\ldots,4, \;\; i = 1,\ldots,27.$$

Or…
$$\begin{aligned}
Y_{i1} &= \beta_0 + \beta_1 Age_{i1} + u_{i1} + u_{i2} Age_{i1} + \varepsilon_{i1} \\
Y_{i2} &= \beta_0 + \beta_1 Age_{i2} + u_{i1} + u_{i2} Age_{i2} + \varepsilon_{i2} \\
Y_{i3} &= \beta_0 + \beta_1 Age_{i3} + u_{i1} + u_{i2} Age_{i3} + \varepsilon_{i3} \\
Y_{i4} &= \beta_0 + \beta_1 Age_{i4} + u_{i1} + u_{i2} Age_{i4} + \varepsilon_{i4}
\end{aligned}$$

Matrix representation: $Y_i = x_i\beta + z_i U_i + \varepsilon_i; \quad i = 1,\ldots,27.$

Add extra column for random slopes

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix}; \quad x_i\beta = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}; \quad z_i U_i = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}; \quad \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}.$$

Note that the random intercept model had a scalar random effect while here we have a vector of dimension (2x1) accounting for intercept and slope.

Matrix form, …+ groups.

$$Y_{ij} = \beta_0 + \beta_1 Age_{ij} + \beta_2 Sex_i + \beta_3 \left(Age \times Sex\right)_{ij} + u_{i1} + u_{i2} Age_{ij} + \varepsilon_{ij};$$

$$\mathrm{j} = 1,...,4; \quad i = 1,...,27;$$

Or…

$$Y_{i1} = \beta_0 + \beta_1 Age_{i1} + \beta_2 Sex_i + \beta_3 \left(Age \times Sex\right)_{i1} + u_{i1} + u_{i2} Age_{i1} + \varepsilon_{i1}$$

$$Y_{i2} = \beta_0 + \beta_1 Age_{i2} + \beta_2 Sex_i + \beta_3 \left(Age \times Sex\right)_{i2} + u_{i1} + u_{i2} Age_{i2} + \varepsilon_{i2}$$

$$Y_{i3} = \beta_0 + \beta_1 Age_{i3} + \beta_2 Sex_i + \beta_3 \left(Age \times Sex\right)_{i3} + u_{i1} + u_{i2} Age_{i3} + \varepsilon_{i3}$$

$$Y_{i4} = \beta_0 + \beta_1 Age_{i4} + \beta_2 Sex_i + \beta_3 \left(Age \times Sex\right)_{i4} + u_{i1} + u_{i2} Age_{i4} + \varepsilon_{i4}$$

Matrix form, …+ groups.

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 Age_{ij} + \beta_2 Sex_i + \beta_3 \left(Age \times Sex\right)_{ij}}_{\text{X matrix}} + \underbrace{u_{i1} + u_{i2} Age_{ij}}_{\text{Z matrix}} + \varepsilon_{ij};$$

$$\text{j} = 1,\dots,4; \quad i = 1,\dots,27;$$

Matrix representation: $Y_i = x_i\beta + z_i U_i + \varepsilon_i; \quad i = 1,\dots,n.$

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix};$$

$$x_i\beta = \begin{bmatrix} 1 & 8 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 12 & 0 & 0 \\ 1 & 14 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \text{ when } Sex_i = 0;$$

$$z_i U_i = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix};$$

$$x_i\beta = \begin{bmatrix} 1 & 8 & 1 & 8 \\ 1 & 10 & 1 & 10 \\ 1 & 12 & 1 & 12 \\ 1 & 14 & 1 & 14 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \text{ when } Sex_i = 1;$$

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}.$$

The General LME model
Matrix notation and dimensions

$$Y_i = x_i \beta + z_i U_i + \varepsilon_i;$$
$$j = 1, \ldots, m_i; \; i = 1, \ldots, n.$$

Y_i is a $(m_i \times 1)$ vector of outcomes from subject i,

β is a $(p \times 1)$ vector of fixed effects,

U_i is a $(q \times 1)$ vector of random effects $u_{i1}, \ldots, u_{iq}$,

x_i is a $(m_i \times p)$ matrix of covariates,

z_i is a $(m_i \times q)$ matrix of covariates, links the random effects to Y_i,

ε_i is a $(m_i \times 1)$ vector or random errors.

Exercise: Verify the values for m_i, n, p, q for each Maxillary Distances models.

The General LME model
Structure of matrices

$$Y_i = x_i\beta + z_i U_i + \varepsilon_i;$$

$$j = 1,\ldots,m_i;\; i = 1,\ldots,n.$$

- The columns of z_i are always a subset of the columns of x_i (that is, every random effect variable must have its fixed counterpart.)

- Fixed and time varying covariates are all included in x_i , examples of the latter are the times of measurement, blood pressure, age, etc.

- Any component of β can be allowed to vary randomly by including the corresponding column of x_i in z_i.

Exercise: verify the matrix composition in the simple models discussed earlier for the Maxillary Distance data.

Covariance Matrix of Y

Random intercept model
Maxillary distance example

$$Y_{ij} = \beta_0 + \beta_1 Age_{ij} + u_{i1} + \varepsilon_{ij}; \quad j = 1,\ldots,4; i = 1,\ldots,27.$$

$$Var(u_{i1}) = \sigma_u^{\,2}$$ Is a scalar.

$$R_i = Var(\varepsilon_{ij}) \equiv \left\{ Var(\varepsilon_{ij}) = \sigma_\varepsilon^{\,2},\ Cov(\varepsilon_{ij}, \varepsilon_{ik}) = 0,\ j = 1,\ldots,4 \right\}_{4\times 4}$$

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix}$$

$$R_i = Cov(\varepsilon_i) = \begin{bmatrix} Var(\varepsilon_{i1}) & Cov(\varepsilon_{i1},\varepsilon_{i2}) & Cov(\varepsilon_{i1},\varepsilon_{i3}) & Cov(\varepsilon_{i1},\varepsilon_{i4}) \\ Cov(\varepsilon_{i1},\varepsilon_{i2}) & Var(\varepsilon_{i2}) & Cov(\varepsilon_{i2},\varepsilon_{i3}) & Cov(\varepsilon_{i2},\varepsilon_{i4}) \\ Cov(\varepsilon_{i1},\varepsilon_{i3}) & Cov(\varepsilon_{i2},\varepsilon_{i3}) & Var(\varepsilon_{i3}) & Cov(\varepsilon_{i3},\varepsilon_{i4}) \\ Cov(\varepsilon_{i1},\varepsilon_{i4}) & Cov(\varepsilon_{i2},\varepsilon_{i4}) & Cov(\varepsilon_{i3},\varepsilon_{i4}) & Var(\varepsilon_{i4},\varepsilon_{i4}) \end{bmatrix}$$

$$= \sigma_\varepsilon^{\,2} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \sigma_\varepsilon^{\,2} I_4 .$$

"Identity matrix"

Note: when referring to the variance of a vector we write "Cov()" while for a scalar we write "Var()".

Covariance Matrix of Y

Random intercept and slope model
Maxillary distance example

$$Y_{ij} = \beta_0 + \beta_1 Age_{ij} + u_{i1} + u_{i2} Age_{ij} + \varepsilon_{ij}; \quad j = 1,\ldots,4; i = 1,\ldots,27.$$

MVN: Natural extension of the univariate normal distribution, from a single response to a vector of responses.

$$U_i = \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix} \sim MVN(0, G_i); \qquad \varepsilon_i \sim MVN(0, R_i).$$

The random vector of dimension (2x1) accounting for intercept and slope has a (2x2) covariance matrix.

$$G_i = Cov(U_i) = \begin{bmatrix} Var(u_{i1}) & Cov(u_{i1}, u_{i2}) \\ Cov(u_{i2}, u_{i1}) & Var(u_{i2}) \end{bmatrix}$$

$$= \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}.$$

$$R_i = \sigma_\varepsilon^{\ 2} I_{m_i}.$$

R_i Is the same as in the random intercept model.

General formula for the Covariance matrix of Y_i

$$Cov(Y_i)_{m_i \times m_i} = z_i G_i z_i^T + R_i = V_i$$

$$Cov(U_i) = Cov\left(\begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}\right) = G_i; \qquad Cov(\varepsilon_i) = Cov\left(\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i1} \\ \varepsilon_{i1} \\ \varepsilon_{i1} \end{bmatrix}\right) = R_i.$$

$$\begin{aligned} Cov(Y_i) &= Cov(x_i\beta + z_i U_i + \varepsilon_i) \\ &= Cov(z_i U_i) + Var(\varepsilon_i) \\ &= z_i G_i z_i^T + R_i \\ &= V_i. \end{aligned}$$

Recall the analogy with the scalar form of the variance of random variables X and Y:
Var(constant)=0
Var(constant X)=constant^2 Var(X)
Var(X+Y)=Var(X) + Var(Y) + 2Cov(X,Y)

General formula for the Variance matrix of Y_i

$$Var(Y_i) = z_i G z_i' + R_i = V_i$$

The variance matrix for the random intercept and slope model

$$V_i = Cov(Y_i) = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{bmatrix} + \sigma^2_{\varepsilon} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The diagonal elements of Vi are:

$$\begin{aligned} Var(Y_{ij}) &= Var(\beta_0 + \beta_1 Age_{ij} + u_{i1} + u_{i2} Age_{ij} + \varepsilon_{ij}) \\ &= g_{11} + 2Age_{ij} g_{12} + Age_{ij}^{\ 2} g_{22} + \sigma_{\varepsilon}^{\ 2}. \end{aligned}$$

The off-diagonal elements of Vi are:

$$Cov(Y_{ij}, Y_{ik}) = g_{11} + (Age_{ij} + Age_{ik}) g_{12} + Age_{ij} Age_{ik} g_{22}.$$

Common Variance – Covariance structures
Within-subjects
(V_i matrix)

Common variance-covariance structures
Three within-individual measurements

$$V_i = Cov(Y_i)$$

Simple: No correlation, constant variance.

Variances the same, covariances=0

$$Var(Y_{ij}) = \sigma^2, Cov(Y_{ij}, Y_{ik}) = 0.$$

$$V_i = \begin{bmatrix} \sigma^2 & 0 & 0 \\ & \sigma^2 & 0 \\ & & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ & 1 & 0 \\ & & 1 \end{bmatrix}$$

Compound symmetry: Constant correlation

Variances the same, covariances the same

$$Var(Y_{ij}) = \sigma_1^{\,2} + \sigma^2, Cov(Y_{ij}, Y_{ik}) = \sigma_1^{\,2}.$$

$$V_i = \begin{bmatrix} \sigma_1^{\,2} + \sigma^2 & \sigma_1^{\,2} & \sigma_1^{\,2} \\ & \sigma_1^{\,2} + \sigma^2 & \sigma_1^{\,2} \\ & & \sigma_1^{\,2} + \sigma^2 \end{bmatrix}$$

I.e., Random intercept model.

Bonate, Table 6.2

Examples of variance-covariance structures
Three within-individual measurements

$$V_i = Cov(Y_i)$$

Unstructured: All variances and covariances are unique to each subject.

Variances different, covariances different

$$Var(Y_{ij}) = \sigma_i^{\ 2}, Cov(Y_{ij}, Y_{ik}) = \sigma_{ik}^{\ 2}.$$

$$V_i = \begin{bmatrix} \sigma_1^{\ 2} & \sigma_{12}^{\ 2} & \sigma_{13}^{\ 2} \\ & \sigma_2^{\ 2} & \sigma_{23}^{\ 2} \\ & & \sigma_3^{\ 2} \end{bmatrix}$$

Toeplitz: Equal variance across measurements, unequal covariances.

Variances the same, covariances different

$$Var(Y_{ij}) = \sigma^2, Cov(Y_{ij}, Y_{ik}) = \sigma_{ik}^{\ 2}$$

$$V_i = \begin{bmatrix} \sigma^2 & \sigma_{12}^{\ 2} & \sigma_{13}^{\ 2} \\ & \sigma^2 & \sigma_{23}^{\ 2} \\ & & \sigma^2 \end{bmatrix}$$

Used with equally time-spaced measurements.

Bonate, Table 6.2

Examples of variance-covariance structures
Three within-individual measurements

First-order autoregressive: Constant variance with decreasing correlation in proportion to the time between measurements

Variances the same, covariances different but with a pattern.

$$Var(Y_{ij}) = \sigma^2, Cov(Y_{ij}, Y_{ik}) = \sigma^2 \rho^{k-j};$$
$$\rho^{k-j} = Corr(Y_{ij}, Y_{ik})$$

$$V_i = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ & \sigma^2 & \sigma^2\rho \\ & & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \\ & 1 & \rho \\ & & 1 \end{bmatrix}$$

Bonate, Table 6.2

4. Estimation Methods and Inference

Specific learning objectives:

1. Explain the estimation methods for LME models.
2. Explain the effect of shrinkage.

What is being estimated?

- Fixed effects $\beta_0, \beta_1, \ldots, \beta_{p-1}$ that is, a vector of size p denoted as β.
- Variance matrices G and R.

E.g. for the random intercept and slope model

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 t_{ij}}_{\text{Fixed effects}} + \underbrace{u_{i1} + u_{i2} t_{ij}}_{\text{Random effects}} + \varepsilon_{ij}; \quad j = 1,\ldots,m_i, i = 1,\ldots,n.$$

3 parameters (g_{11}, g_{12}, g_{22})

Parameters to estimate: 2 fixed effect parameters
+ random effects variance parameters: G matrix and
+ random residuals variance: R matrix

1 parameter (σ_ε^2)

$\boxed{\text{Total = 6 parameters}}$

Estimation in LME models

Steps:

1. Estimate variance components G and R via ***Maximum Likelihood (ML) methods***

2. Estimate fixed and random effects via ***Generalized Least Squares (GLS)***

What ML methods are available?
What is GLS?

Generalized Least Squares
GLS

- OLS (Ordinary Least Squares) no longer applicable since errors are not independent in mixed effects models.
- The goal of GLS is to minimize a weighted version of the error squares.

Objective function

OLS goal is to minimize:

$$(Y - X\beta)'(Y - X\beta)$$

Matrix representation of the sum

$$\varepsilon_1^{\ 2} + \varepsilon_2^{\ 2} + \cdots + \varepsilon_n^{\ 2}$$

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i.$$

GLS goal is to minimize:

$$(Y - X\beta)'V^{-1}(Y - X\beta)$$

V=Var(Y) involves the G and R Matrices, already estimated via ML methods.

Maximum Likelihood Based Estimation

- Requires distributional assumption for the data.

- Normality assumption of the random effects and the residual variability imposes a Multivariate Normal Distribution for the response.

ML Estimation Methods for V_i

ML vs. Restricted ML (REML)

- ML procedure: simultaneously estimates fixed effects (β) and variance components (G, R) by maximizing the likelihood function for a Multivariate Normal Distribution.

- REML procedure: maximizes a function that is *only* a function of the variance components (G,R) and no fixed effects.

"Basically, the main idea is to separate that part of the data used for estimation of V_i from that used for estimation of β. Estimation of V_i is based only on the relevant part of the data." (Fitzmaurice, Laird & Ware, 2011).

Estimation Procedure

1. Estimates of G and R are first obtained via ML or REML methods.

2. Estimates for the fixed effects β are obtained by solving the GLS equations and "plugging in" the estimated G and R obtained above.

3. Standard errors of estimates for the fixed effects parameters can be obtained from the estimated variance-covariance matrix V.

Estimates for fixed and random effects:

$$\hat{\beta}_{GLS} = \left(X^T \hat{V}^{-1} X\right)^{-1} X^T \hat{V}^{-1} Y$$

"Weighted Least Squares Estimate"

Recall, for OLS:

$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T Y$$

Estimation Procedure

- When estimating G and R, REML is less biased than ML for small sample sizes.

- Think of sample SD vs. ML SD: recall that ML SD is biased for small sample sizes. A similar principle applies with G and R as REML adjusts for the degrees of freedom.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}{n-1} \quad \text{vs.} \quad \hat{\sigma}_{ML}{}^2 = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}{n}$$

- The difference between REML and ML estimates for G and R becomes less important as the number of subjects in the sample becomes larger than the number of fixed parameters in the model.

What about the random effects?

- Random effects (u_i's) are often said to be "predicted" rather than "estimated".

- This is because they are random variables and not fixed population parameters.

- Prediction of the random effects can be done once the estimation of the fixed effects and covariance of Y is done.

Prediction of Random Effects U_i

- By noting that U_i and Y_i have a joint Multivariate Normal Distribution, it can be shown that

$$E\left(U_i \mid Y_i\right) = G_i z_i^T V_i^{-1}\left(Y_i - X_i\hat{\beta}\right)$$

Based on the conditional probability of U|Y and the use of the Bayes Theorem in probability.

E.g. In the random intercept & slope model:

$$U_i = \begin{bmatrix} u_{i1} \\ u_{i2} \end{bmatrix}$$

- It is valid to “plug in” the estimates of G and V above:

$$\hat{U}_i = \hat{G}_i z_i^T \hat{V}_i^{-1}\left(Y_i - X_i\hat{\beta}\right)$$

This expression is called:
“Empirical Best Linear Unbiased Predictor (BLUP)” or
“Emprical Bayes Estimate” (EBE).

Shrinkage
Empirical Bayes Estimate (EBE)

- Compromise between what is observed and the population response:
  - BSV>>>WSV, individual predicted values get closer to individual observed response.
  - WSV>>>BSV, individual predicted values get closer to population mean response.
  - Pulls extreme observations to population mean, when WSV is very high.

Shrinkage
Empirical Bayes Estimate (EBE)

- The fitted or predicted value of Y_i can be re-expressed in terms of the covariance matrices R_i and $V_i = G_i + R_i$:

$$\hat{Y}_i = X_i \hat{\beta} + z_i \hat{U}_i$$

$$\hat{Y}_i = \left[\hat{R}_i \hat{V}_i^{-1}\right] X_i \hat{\beta} + \left[I_{m_i} - \hat{R}_i \hat{V}_i^{-1}\right] Y_i$$

- Recall G_i is BSV and R_i is WSV.

BSV>>WSV	$\left[\hat{R}_i \hat{V}_i^{-1}\right] \approx "0"$	$\leftrightarrow$	$\hat{Y}_i \approx Y_i$
WSV>>BSV	$\left[\hat{R}_i \hat{V}_i^{-1}\right] \approx \mathrm{I}_{\mathrm{m}_i}$	$\leftrightarrow$	$\hat{Y}_i \approx X\hat{\beta}$

Inference

Specific learning objectives:

1. Implement hypothesis tests for the fixed effects via conditional t-tests and F-tests in R.
2. Implement LRT for hypothesis tests of fixed effects and covariance in R.

Likelihood Ratio Test (LRT)

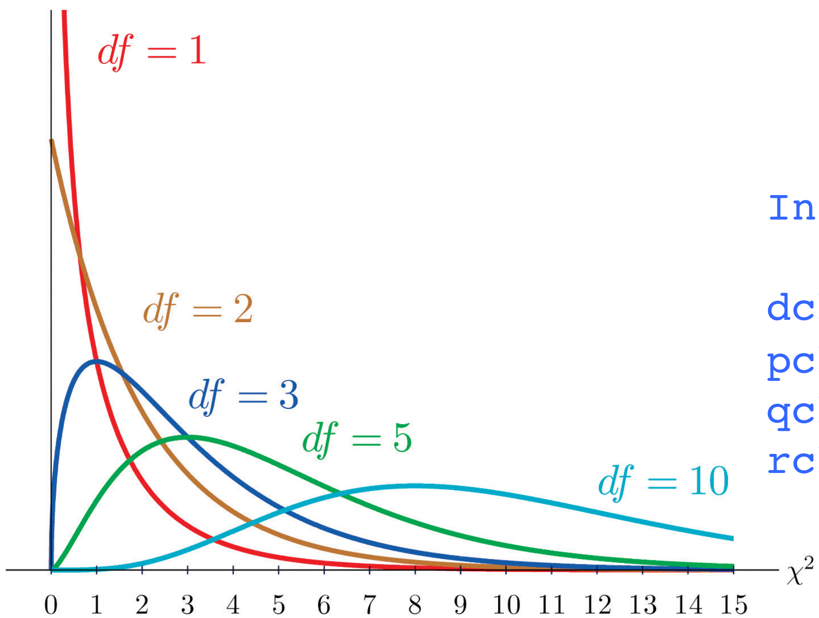
Similar idea of the ANOVA F-test for nested models.

If L_2 is the likelihood of the unrestricted model with k_2 parameters and L_1 is the likelihood of the restricted model with k_1 parameters, $k_1<k_2$.

The LRT test statistic is

$$2\ \log(L_2/L_1) = 2\left[\log L_2 - \log L_1\right] \ \sim \chi^2(k_2-k_1)$$

Chi-squared Distribution with k2-k1 degrees of freedom.




```
In R:

dchisq(quantile, df)      density function
pchisq(quantile, df)      probability
qchisq(probability, df)   quantile
rchisq(n, df)             random generation
```

Inference for random effects
Likelihood Ratio Test (LRT)

The relationship between the REML and the ML functions is given below:

$$\log L^{REML} = \log L^{ML} - \frac{1}{2}\log\left|\sum_{i=1}^{n} X_i^T V_i^{-1} X_i\right|$$

The LRT with REML can be used if :

- Both models were fit using REML and
- The fixed effects specification is the same for both models.

It is recommended to

- Use the LRT with REML when comparing nested models for the covariance.
- Use the LRT with ML when comparing nested models for the fixed effects.
- LRT with REML is not to be used when testing fixed effects as L2 and L1 differ due to the extra term above.

Fitzmaurice, Laird & Ware, 2011.

Hypothesis tests for fixed effects

Wald tests: since we have assumed a Normal model for the Y's (through normality of the random effects and residuals), the following can be used:

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim N(0,1)$$

Good for large samples

So hypothesis tests of the form $H_0{:}\beta_j=0$ can be performed as usual.

Problem with Wald tests:

- The variance of β involves elements of the matrices Vi = Gi + Ri and no correction is made above for the uncertainty of the estimates of Gi and Ri.
- There is a risk of the variance to be underestimated as it does not consider the extra variability that the estimates of Gi and Ri bring.

Inference for fixed effects

Alternative 1:

- t-distribution the test statistic instead of N(0,1).

Alternative 2:

- Likelihood Ratio Tests (to be seen shortly)

Alternative 3:

- Some authors strictly recommend not to use LRT tests to assess fixed effects (Pinheiro and Bates), arguing these are too liberal (p-values are too small).

- They suggest the use of so called conditional t and F tests such as those implemented in the R `summary()` function.

Other likelihood based measures to compare models

- Akaike Information Criterion

$$AIC = -2 \log L^{REML} + 2 \text{ cov.par}$$

  - Where cov.par is the number of covariance parameters.
  - Used to compare non-nested models for the covariance that have the same fixed effects.
  - One should select the model that minimizes AIC.

- Bayes Information Criterion

$$BIC = - 2 \log L + \text{cov.par} \log(n)$$

  - Where n is the number of subjects and cov.par as before.
  - The model that minimizes BIC is preferred.

AIC is preferred over BIC for covariance selection, see Fitzmaurice for details.

Example: Maxillary Distance Data (Orthodont)
Testing for the significance of the fixed effects

When to use REML or ML for LRT to test fixed effects

- Recall that for the Maxillary data example we fitted
  Random intercept & slope + sex
  Random intercept & slope + sex + sex:age

The R output can be used to test for the individual significance of the estimates for sex, age and sex:age.

However, these two models cannot be used in a LRT to test for a joint hypothesis, for example, sex + sex:age.

This is because the lme() function has used “REML” by default. In order to test for fixed effects via LRT, we must specify method=“ML” in the lme() function.

Example: Maxillary Distance Data (Orthodont)

Restricted model, L_1
"fitt.sex.ml" for random intercept & slope + sex

Unrestricted model, L_2
"fitt.sexage.ml" for random intercept & slope + sex + age:sex

The LRT is done through the `anova()` function

```
> fitt.sex.ml <- lme(distance~I(age-11)+Sex,
                   data=dat,random=~I(age-11)|Subject,method="ML")
> fitt.sexage.ml <- lme(distance~I(age-11)+Sex+I(age-11):Sex,
                   data=dat,random=~I(age-11)|Subject,method="ML")
>
> anova(fitt.sex.ml,fitt.sexage.ml)
```

```
               Model df      AIC      BIC    logLik   Test L.Ratio p-value
fitt.sex.ml        1  7 446.8352 465.6101 -216.4176
fitt.sexage.ml     2  8 443.8060 465.2630 -213.9030 1 vs 2 5.02921  0.0249
```

Based on this p-value, we conclude that the interaction term is significant at a 5% level.

Note that the AIC and BIC favor the model with interaction term.

```
> fitt.sex <- update(fitt2,.~.+Sex)
> summary(fitt.sex)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  449.2339 467.8116 -217.6169

Random effects:
 Formula: ~I(age - 11) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 1.8320242 (Intr)
I(age - 11) 0.2264279 0.19
Residual    1.3100396

Fixed effects: distance ~ I(age - 11) + Sex
                Value Std.Error DF  t-value p-value
(Intercept) 24.897236 0.4852090 80 51.31239   0.000
I(age - 11)  0.660185 0.0712533 80  9.26533   0.000
SexFemale   -2.145489 0.7574536 25 -2.83250   0.009
 Correlation:
            (Intr) I(-11)
I(age - 11)  0.085
SexFemale   -0.636  0.000

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.08141614 -0.45675578  0.01552687  0.44704106  3.89437718

Number of Observations: 108
Number of Groups: 27
```

```
> fitt.sexage <- update(fitt.sex,.~.+Sex:I(age-11))
> summary(fitt.sexage)
Linear mixed-effects model fit by REML
 Data: dat
      AIC      BIC    logLik
  448.5817 469.7368 -216.2908

Random effects:
 Formula: ~I(age - 11) | Subject
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 1.8303267 (Intr)
I(age - 11) 0.1803454 0.206
Residual    1.3100397

Fixed effects: distance ~ I(age - 11) + Sex + Sex * I(age - 11)
                          Value Std.Error DF  t-value p-value
(Intercept)           24.968750 0.4860007 79 51.37596  0.0000
I(age - 11)            0.784375 0.0859995 79  9.12069  0.0000
SexFemale             -2.321023 0.7614168 25 -3.04829  0.0054
I(age - 11):SexFemale -0.304830 0.1347353 79 -2.26243  0.0264
 Correlation:
                      (Intr) I(g-11) SexFml
I(age - 11)            0.102
SexFemale             -0.638 -0.065
I(age - 11):SexFemale -0.065 -0.638   0.102

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-3.168078484 -0.385939134  0.007103929  0.445154686  3.849463230

Number of Observations: 108
Number of Groups: 27
```

Example: Maxillary Distance Data (Orthodont)

Testing for the significance of the random slope

The LRT with REML must be used to assess the significance of the covariance elements.

Recall our models:
"fitt1" for random intercept -> restricted model, L_1
"fitt2" for random intercept & slope -> unrestricted model, L_2

The LRT is done through the `anova()` function:

Preferred for being REML

```
> anova(fitt1,fitt2)
      Model df      AIC      BIC    logLik   Test L.Ratio p-value
fitt1     1  4 455.0025 465.6563 -223.5013
fitt2     2  6 454.6367 470.6173 -221.3183 1 vs 2 4.36583  0.1127

> fitt1.ML <- update(fitt1,method="ML")
> fitt2.ML <- update(fitt2,method="ML")

> anova(fitt1.ML,fitt2.ML)
      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
fitt1     1  4 451.3895 462.1181 -221.6948
fitt2     2  6 451.2116 467.3044 -219.6058 1 vs 2 4.177941  0.1238
```

We conclude that the model with the random slope is not better than the model with random intercept.

Effect of Centering Age

- Centering Age does not only make sense for the interpretation of the intercept, it also makes the estimates for β_0 and β_1 uncorrelated.
- Applicability of this is a simplification of the calculation of CI for marginal mean.

In the simplest case with no grouping factor:

$$E(Y_{ij}) = \mu_{ij} = \beta_0 + \beta_1 Age^*_{ij}$$

$$\hat{E}(Y_{ij}) = \hat{\mu}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Age^*_{ij}$$

$$Var(\hat{\mu}_{ij}) = Var(\hat{\beta}_0) + {Age^*_{ij}}^2 Var(\hat{\beta}_1) + 2Age^*_{ij} Cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$\hat{Var}(\hat{\mu}_{ij}) = \hat{Var}(\hat{\beta}_0) + {Age^*_{ij}}^2 \hat{Var}(\hat{\beta}_1) + 2Age^*_{ij} \hat{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$= 0.429^2 + {Age^*_{ij}}^2 0.062^2 + 0$$

$$\hat{Var}(\hat{\mu}_{i1}) = 0.429^2 + 8^2 0.062^2 = 28.079$$

$$95\%\ CI \text{ for } \hat{\mu}_{i1}: \quad \left(\hat{\mu}_{i1} \pm 1.96 SE(\hat{\mu}_{i1})\right) = \left(29.30463 \pm 1.96\sqrt{28.079}\right)$$

$$= (18.919, 39.691)$$

Effect of Centering Age

- Centering Age does not only make sense for the interpretation of the intercept, it also makes the estimates for β_0 and β_1 uncorrelated.
- Applicability of this is a simplification of the calculation of CI for marginal mean.

In the simplest case with no grouping factor:

$$E\left(Y_{ij}\right) = \mu_{ij} = \beta_0 + \beta_1 Age^*_{ij}$$

$$\hat{E}\left(Y_{ij}\right) = \hat{\mu}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Age^*_{ij}$$

$$Var(\hat{\mu}_{ij}) = Var(\hat{\beta}_0) + Age^{*\,2}_{ij} Var(\hat{\beta}_1) + 2Age^*_{ij} Cov(\hat{\beta}_0, \hat{\beta}_1)$$

$$\hat{Var}(\hat{\mu}_{ij}) = \hat{Var}(\hat{\beta}_0) + Age^{*\,2}_{ij} \hat{Var}(\hat{\beta}_1) + 2Age^*_{ij} \hat{Cov}(\hat{\beta}_0, \hat{\beta}_1)$$

$$= 0.429^2 + Age^{*\,2}_{ij} 0.062^2 + 0$$

$$\hat{Var}(\hat{\mu}_{i1}) = 0.429^2 + 8^2 0.062^2 = 28.079$$

$$95\%\ CI \text{ for } \hat{\mu}_{i1}: \quad \left(\hat{\mu}_{i1} \pm 1.96 SE(\hat{\mu}_{i1})\right) = \left(29.30463 \pm 1.96\sqrt{28.079}\right)$$

$$= \left(18.919, 39.691\right)$$

```
> summary(fitt1)
Linear mixed-effects model fit by REML
 Data: dat
       AIC      BIC    logLik
  455.0025 465.6563 -223.5013

Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    2.114724 1.431592

Fixed effects: distance ~ I(age - 11)
               Value Std.Error DF  t-value p-value
(Intercept) 24.023148 0.4296605 80 55.91193       0
I(age - 11)  0.660185 0.0616059 80 10.71626       0
 Correlation:
            (Intr)
I(age - 11) 0

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.66453932 -0.53507984 -0.01289591  0.48742859  3.72178465

Number of Observations: 108
Number of Groups: 27
```

Correlation of fixed effects estimates