Lecture 7 | 07.04.2025

Linear mixed effects model
(some theoretical and empirical issues)

A brief overview

- Simple linear regression model for repeated measurements within $N \in \mathbb{N}$ (independent) subjects ($i \in \{1, \ldots, N\}$) where

$$\boldsymbol{Y}_i = \mathbb{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

  for the response vector $\boldsymbol{Y}_i \in \mathbb{R}^{n_i}$ where $\mathbb{X}_i = (\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i})^\top$, $\boldsymbol{X}_{ij} \in \mathbb{R}^p$ for $j = 1, \ldots, n_i$ are the explanatory vectors—measurements at time points $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^\top$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the unknown vector of parameters

- The variance-covariance structure within each subject is modelled by the vector parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$, such that $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\boldsymbol{0}_i, \mathbb{V}_i(\boldsymbol{t}_i, \boldsymbol{\alpha}))$, where

$$\varepsilon_{ij} = \boldsymbol{z}_{ij}^\top \boldsymbol{w}_i + W_i(t_{ij}) + \omega_{ij}$$

  for random vector $\boldsymbol{w}_i$, random process $W_i(t)$, and random variable ω_{ij}

- This can be rewritten as a **linear mixed (effects) model (LMM)** with fixed effects β, random effects $\boldsymbol{w}_i$'s, and the error terms $\boldsymbol{R}_i$'s

$$\boldsymbol{Y}_i = \mathbb{X}_i \boldsymbol{\beta} + \mathbb{Z}_i \boldsymbol{w}_i + \boldsymbol{R}_i,$$

  where $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{in_i})^\top = (W_i(t_{i1}) + \omega_{i1}, \ldots, W_i(t_{in_i}) + \omega_{in_1})^\top$
  (different formulations of the same model depending on which part of the model is emphasized)

Two stage approach vs. LMM formulation

- Considering the longitudinal data $\{(Y_{ij}, \boldsymbol{X}_{ij});\ i = 1, \dots, N; j = 1, \dots, n_i\}$ the statistical analysis can be either performed in a two stage process
  - (1) separate models $\boldsymbol{Y}_i = \mathbb{X}_i^{(1)}\beta_i + \varepsilon_i$ for each subject $i = 1, \dots, N$
  - (2) and the overall model for regression parameters $\beta_i = \mathbb{X}_i^{(2)}\beta + \boldsymbol{b}_i$

- Alternatively (but not equivalently), one common model with mixed effects (LMM) can be used instead where

$$\left.\begin{array}{l} \boldsymbol{Y}_i = \mathbb{X}_i^{(1)}\beta_i + \varepsilon_i \\ \beta_i = \mathbb{X}_i^{(2)}\beta + \boldsymbol{b}_i \end{array}\right\} \Longrightarrow \quad \boldsymbol{Y}_i = \underbrace{\mathbb{X}_i^{(1)}\mathbb{X}_i^{(2)}}_{\mathbb{X}_i}\beta + \underbrace{\mathbb{X}_i^{(1)}}_{\mathbb{Z}_i}\underbrace{\boldsymbol{b}_i}_{\boldsymbol{w}_i} + \underbrace{\varepsilon_i}_{\boldsymbol{R}_i}$$

What are common drawbacks of the two-stage model formulation that are overcome in the overall LMM formulation?

Consider, for instance, a linear regression line in the first stage and a subject with only one observations. Or, instead, a quadratic fit in the first stage and some subjects with only two measurements?

Components of the LMM

- **Fixed effects $\mathbb{X}_i\boldsymbol{\beta}$**
  - the same structure for all subjects (the population mean structure)
  - covariates $\boldsymbol{X}_{ij}$ are generally assumed to be random but the regression framework is typically considered conditionally on the model matrix $\mathbb{X}$

- **Random effects $\mathbb{Z}_i\boldsymbol{w}_i$**
  - the subject-specific part of the model (the individual mean structure)
  - describes how the mean parameters for one subject differ from the mean parameters for the other subject—resp. how the population mean (common) differs from the subject's specific mean (individual)

- **Non-systematic terms (error) $\boldsymbol{R}_i$**
  - sometimes called the variance components model
  - accounts for the between and withing subjects' variability
  - partially modeled by the subject specific covariates...
    (typically when heterogeneity is modeled withing specific groups)

Population vs. individual interpretation

Consider LMM of the form $\boldsymbol{Y}_i = \mathbb{X}_i\boldsymbol{\beta} + \mathbb{Z}_i\boldsymbol{w}_i + \boldsymbol{R}_i$ where, typically, $\boldsymbol{w}_i \sim N(0, \mathbb{G})$ and $\boldsymbol{R}_i \sim N(\boldsymbol{0}, \mathbb{R}_i)$ – alternatively $\boldsymbol{Y} = \mathbb{X}\boldsymbol{\beta} + \mathbb{Z}\boldsymbol{w} + \boldsymbol{R}$

- Marginal model $\boldsymbol{Y}_i \sim N(\mathbb{X}_i\boldsymbol{\beta}, \mathbb{Z}_i\mathbb{G}\mathbb{Z}_i^\top + \mathbb{R}_i)$

  A population characterization and a population interpretation of the model—the model describes the conditional mean given a subset of specific (sub-population) characteristics. Inference with respect to the subpopulation differences

- Hierarchical model $\boldsymbol{Y}_i|\boldsymbol{w}_i \sim N(\mathbb{X}_i\boldsymbol{\beta} + \mathbb{Z}_i\boldsymbol{w}_i, \mathbb{R}_i)$ and $\boldsymbol{w}_i \sim N(0, \mathbb{G})$

  Subject specific characterization and subject specific as well as population interpretation of the model—the model describes—in two levels (therefore hierarchical)—the conditional mean of a specific subject i but it can be integrated over the distribution of $\boldsymbol{w}_i$ to obtain the population characterization (similarly as in the marginal model)

$\hookrightarrow$ note, that the hierarchical model can be used to obtain the marginal model, but this does not hold in vise-versa manner. Also, different hierarchical models can produce the same marginal model

Examples

- **Example 1** Consider a simple linear mixed effect model for two repeated observations only (i.e., $n_i = 2$) with a random intercept term and uncorrelated heterogenous errors $\boldsymbol{R}_i = (R_{i1}, R_{i2})^\top$ where $R_{i1} \sim N(0, \tau_1^2)$ and $R_{i2} \sim N(0, \tau_2^2)$. What is the mean structure? What is the overall variance-covariance structure $\mathbb{Z}_i \mathbb{G} \mathbb{Z}_i^\top + \mathbb{R}_i$?

- **Example 2** Consider a simple linear mixed effect model for two repeated observations only (i.e., $n_i = 2$) with (uncorrelated) random intercept and random slope terms and homoscedastic errors $\boldsymbol{R}_i \sim N_2(\mathbf{0}, \tau_2 \mathbb{I})$. What is the mean structure? What is the overall variance-covariance structure?

Thus, as a direct consequence, any good marginal model fit can not be used as an argument to justify also a good hierarchical model fit...

We can only contradict a wrong model... we can not prove a right model!

Inference in a marginal model

Basically, there are two parts of the model that are important when performing the statistical inference about the unknown parameters

- Inference about the fixed effects (parameters $\boldsymbol{\beta} \in \mathbb{R}^p$)
  - Wald type tests
  - t-tests and F-tests
  - likelihood ratio tests
  - robust (sandwich) inference

- Inference about variance/covariance components (parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$)
  - Wald type tests
  - likelihood ratio tests

$\hookrightarrow$ in practical applications there are also various information criteria used (AIC, BIC, Hannan and Quinn (HQIC), Bozdogan (CAIC), etc.)
(log-likelihood minus penalty: $\#\theta \mid (\#\theta \log \mathcal{N})/2) \mid \#\theta \log\log \mathcal{N} \mid \#\theta(\log \mathcal{N} + 1)/2$,
where $\mathcal{N} = \sum_i n_i$ for ML and $\mathcal{N} = N - p$ for REML – but ML should be used for comparisons)

Inference for the mean structure (I)

- the estimate for $\boldsymbol{\beta} \in \mathbb{R}^p$

$$\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}) = \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1} \mathbb{X}^\top \mathbb{W} \boldsymbol{Y}$$

where $\mathbb{W}^{-1} = \mathbb{V}(\boldsymbol{\alpha}, \boldsymbol{t})$, and $\widehat{\boldsymbol{\alpha}} \in \mathbb{R}^q$ is a REML (ML) estimate of $\boldsymbol{\alpha} \in \mathbb{R}^q$ *($\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$ is unbiased estimate whatever the value of $\widehat{\boldsymbol{\alpha}} \in \mathbb{R}^q$ is plugged-in)*

- the variance of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$ is

$$Var[\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})] = \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1} \Big(\mathbb{X}^\top \mathbb{W}^\top [Var\,\boldsymbol{Y}] \mathbb{W} \mathbb{X}\Big) \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1}$$

and for a correctly specified variance matrix $Var[\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})] = \left(\mathbb{X}^\top \mathbb{W} \mathbb{X}\right)^{-1}$

- the distribution of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}})$ is (conditionally on $\widehat{\boldsymbol{\alpha}}$) approximately normal, with the corresponding mean and variance structure

Inference for the mean structure (I)

Consider the null hypothesis of the form $H_0 : \mathbb{L}\beta = \mathbf{0}$ vs. $H_A : \mathbb{L}\beta \neq \mathbf{0}$

- Wald tests (approximate)

$$T = \widehat{\boldsymbol{\beta}}^\top \mathbb{L}^\top \left[\mathbb{L}\Big(\mathbb{X}^\top \mathbb{V}^{-1}(\boldsymbol{t}, \widehat{\boldsymbol{\alpha}})\mathbb{X}\Big)^{-1}\mathbb{L}^\top\right]^{-1} \mathbb{L}\widehat{\boldsymbol{\beta}} \quad \underset{as.}{\overset{H_0}{\sim}} \quad \chi^2_{rank(\mathbb{L})}$$

(but the additional variability introduced by replacing $\boldsymbol{\alpha}$ with $\widehat{\boldsymbol{\alpha}}$ is not accounted for)

- t-tests and F-tests (approximate)

$$F = \frac{\widehat{\boldsymbol{\beta}}^\top \mathbb{L}^\top \left[\mathbb{L}\Big(\mathbb{X}^\top \mathbb{V}^{-1}(\boldsymbol{t}, \widehat{\boldsymbol{\alpha}})\mathbb{X}\Big)^{-1}\mathbb{L}^\top\right]^{-1} \mathbb{L}\widehat{\boldsymbol{\beta}}}{rank(\mathbb{L})} \quad \underset{as.}{\overset{H_0}{\sim}} \quad F_{rank(\mathbb{L}),M}$$

(where M needs to be approximated—containment, Satterthwaite, or Kenward & Roger)

- Likelihood ratio tests (approximate)

$$L = -2\ln\lambda = -2\ln\Big[L(model\ H_0)/L(model\ H_A)\Big] \quad \underset{as.}{\overset{H_0}{\sim}} \quad \chi^2_{dim(H_A)-dim(H_0)}$$

Inference for the variance structure (II)

Both, ML and REML estimates of $\boldsymbol{\alpha} \in \mathbb{R}^q$ are approximately normally distributed with the true value as the mean vector and the inverse Fisher information matrix as the variance-covariance matrix

- **Approximate Wald type tests** are easily obtainable
  (in SAS the option `covtest` *in the* `proc mixed` *statement)*
- However, some statistical tests may not have any reasonable interpretation under the hierarchical model (the tests are only meaningful under the marginal model) (Consider: $\mathit{Var}\,\boldsymbol{Y}_i(t)) = (1, t)\mathbb{G}(1, t)^\top + \sigma^2)$
- Moreover, the quality of the normal approximation depends on the true value of $\boldsymbol{\alpha} \in \mathbb{R}^q$ and the approximation completely fails when testing for boundary values (Again marginal vs. hierarchical model)
- **Likelihood ratio tests** for comparisons of nested models
  (also valid for REML if the same mean structure is used)

Inference for the variance structure (II)

- Under the hierarchical model the asymptotic distribution of the likelihood ratio test statistic under the null hypothesis of variance component insignificance can be derived
- For a model with random effects (intercept and slope) $\boldsymbol{w}_i \sim N_2(\mathbf{0}, \mathbb{G})$ the significance of the random slope can be tested by the null hypothesis

$$H_0 : g_{12} = g_{21} = g_{22}$$

- Note, that the null hypothesis (specifically $g_{22} = 0$, where $\mathbb{G} = (g_{ij})_{i,j=1}^{2}$ is actually on the boundary of the parameter space—therefore, a normal approximation for the the corresponding estimate $\widehat{g}_{22}$ is not appropriate
- It can be show, that under the null hypothesis H_0 the likelihood ratio test statistic $-2\ln\lambda_N$ converges in distribution to a mixture of two χ^2 distributions:

$$-2\ln\lambda_N \xrightarrow{\mathcal{D}} \chi^2_{1:2}, \quad \text{for } N \to \infty$$

Inference on individual profiles (III)

The measurements of the dependent variable, $Y_{ij} \in \mathbb{R}$, for subjects $i = 1, \dots, N$ and repeated observations $j = 1, \dots, n_i$ within the subject i (taken at the time-points $t_{i1} < t_{i2} < \cdots < t_{in_i}$) can be also expressed as

$$Y_i(t_{ij}) \equiv Y_{ij} = \mu(t_{ij}) + U_{ij} + W_i(t_{ij}) + \omega_{ij}$$

where

- $\mu(t_{ij}) \equiv \boldsymbol{X}_{ij}^\top \beta$ is the mean profile
- $U_{ij} = \boldsymbol{z}_{ij}^\top \boldsymbol{w}_i$, where $U_{ij} \sim N(0, \boldsymbol{z}_{ij}^\top \mathbb{G} \boldsymbol{z}_{ij})$, independent in $i \in \{1, \dots, N\}$
- $W_i(t_{ij})$ are realization of independent copies $\{W_i(t)\}$ of a zero mean Gaussian process with the covariance function $\sigma^2 \rho(u)$
- $\omega_{ij} \sim N(0, \tau^2)$ are mutually independent measurement errors

Goal: To construct an estimate (a prediction) for an individual i outcome at the time point t, meaning that we want to obtain the profile for $\widehat{Y}_i(t)$

Towards the individual's prediction

- as far as $\omega_{ij} \sim N(0, \tau^2)$ are zero-mean (independent) measurement errors they do not contribute to the prediction/estimation of $Y_i(t)$
- therefore, the prediction/estimate of $Y_i(t)$ can be expressed as

$$\widehat{Y}_i(t) = \widehat{\mu}(t) + \widehat{U} + \widehat{W}_i(t) = \widehat{\mu}(t) + \widehat{\Omega}_i(t)$$

where $\widehat{\mu}(t)$ represents the estimate for the mean structure and $\widehat{\Omega}_i(t)$ represents the estimate for the variance/covariance structure

  - the mean structure can be estimated by standard techniques (e.g., by assuming a linear regression model)
  - **How to estimate the variance/covariance structure $\Omega_i(t)$?**

Continuous process vs. discrete realizations

- the subject specific profile $Y_i(t)$ is only observed at some finite number of time points $\boldsymbol{t}_i = (t_{i1}, \dots, t_{in_i})^\top \in \mathbb{R}^{n_i}$
- the same can be also said about the subject specific variance/covariance structure $\Omega_i(t)$ that is only observed at $\boldsymbol{t}_i = (t_{i1}, \dots, t_{in_i})^\top \in \mathbb{R}^{n_i}$
- Analogously, the estimate for $\Omega_i(t)$ will be only provided for some specific (finitely many) time points, lets say $\boldsymbol{t} = (t_1, \dots, t_n)^\top \in \mathbb{R}^n$
- Under the assumed normality, we have $\boldsymbol{Y}_i \sim N(\mathbb{X}_i\boldsymbol{\beta}, \mathbb{Z}_i\mathbb{G}\mathbb{Z}_i^\top + \sigma^2\mathbb{H}_i + \tau^2\mathbb{I}_i)$ and also $\boldsymbol{\Omega}_i = (\Omega_i(t_1), \dots, \Omega_i(t_n))^\top \sim N(\boldsymbol{0}, \mathbb{Z}_t\mathbb{G}\mathbb{Z}_t^\top + \sigma^2\mathbb{H}_t)$ where $\mathbb{Z}_t$ and $\mathbb{H}_t$ correspond to the time points $\boldsymbol{t} = (t_1, \dots, t_n)^\top$
- Thus, it also holds that

$$\begin{pmatrix} \boldsymbol{\Omega}_i \\ \boldsymbol{Y}_i \end{pmatrix} \sim N_{n+n_i}\left(\begin{pmatrix} \boldsymbol{0} \\ \mathbb{X}_i\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}(\boldsymbol{t},\boldsymbol{t}) & \boldsymbol{\Sigma}(\boldsymbol{t},\boldsymbol{t}_i) \\ \boldsymbol{\Sigma}(\boldsymbol{t}_i,\boldsymbol{t}) & \tau^2\mathbb{I} + \boldsymbol{\Sigma}(\boldsymbol{t}_i,\boldsymbol{t}_i) \end{pmatrix}\right)$$

where $\boldsymbol{\Sigma}(\cdot,\cdot)$ represent the corresponding covariance matrix

Conditional normal distribution

- **A natural estimate for $\boldsymbol{\Omega}_i$ would the the conditional expectation, i.e.**

$$\widehat{\boldsymbol{\Omega}}_i = E[\boldsymbol{\Omega}_i | \boldsymbol{Y}_i]$$

- Using standard properties of a multivariate normal distribution, where

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \sim N_{p+q}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

  it holds that

  - the conditional expectation of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is

$$E[\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}] = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_2) \qquad (\boldsymbol{\mu}_{X|Y})$$

  - the conditional variance of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is

$$Var[\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \qquad (\Sigma_{X|Y})$$

  - thus, the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$ is

$$\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y} \sim N_p(\mu_{X|Y}, \Sigma_{X|Y})$$

Estimate for the subject's profile

Using now the properties of the multivariate normal distribution we finally obtain

- $\widehat{\boldsymbol{\Omega}}_i = E[\boldsymbol{\Omega}_i | \boldsymbol{Y}_i] = \boldsymbol{\Sigma}(\boldsymbol{t}, \boldsymbol{t}_i)\Big[\tau^2\mathbb{I} + \boldsymbol{\Sigma}(\boldsymbol{t}_i, \boldsymbol{t}_i)\Big]^{-1}(\boldsymbol{Y}_i - \mathbb{X}_i\boldsymbol{\beta})$
- $Var[\widehat{\boldsymbol{\Omega}}_i | \boldsymbol{Y}_i] = \boldsymbol{\Sigma}(\boldsymbol{t}, \boldsymbol{t}) - \boldsymbol{\Sigma}(\boldsymbol{t}, \boldsymbol{t}_i)\Big[\tau^2\mathbb{I} + \boldsymbol{\Sigma}(\boldsymbol{t}_i, \boldsymbol{t}_i)\Big]^{-1}\boldsymbol{\Sigma}(\boldsymbol{t}_i, \boldsymbol{t})$

- in the expressions above there are still some quantities that are unknown (the vector of the regression parameters $\boldsymbol{\beta} \in \mathbb{R}^p$ or the parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$ that specifies the variance/covariance structure)
- plug-in techniques are typically used to obtain the final estimate for $\boldsymbol{\Omega}_i$
- note, that for $\tau^2 = 0$ and $\boldsymbol{t} \equiv \boldsymbol{t}_i$, the estimator/predictor $\widehat{\boldsymbol{\Omega}}_i$ reduces to $(\boldsymbol{Y}_i - \mathbb{X}_i\widehat{\boldsymbol{\beta}})$ with zero variance (meaning that if there is no measurement error than the data are perfect estimate/prediction for the true outcome at the existing observation time points)
- when $\tau^2 > 0$, than $\widehat{\boldsymbol{\Omega}}_i$ reflects some compromise between $(\boldsymbol{Y}_i - \mathbb{X}_i\widehat{\boldsymbol{\beta}})$ and zero tending to zero when τ^2 increases

Examples

- **Example 1** Assume a simple linear (regression) model with a random intercept term (i.e., $\boldsymbol{z}_{ij} = 1$ and $w_i \sim N(0, \nu^2)$)
  - observations $Y_i(t_{ij}) \equiv Y_{ij} = \mu(t_{ij}) + U_i + W_i(t)_{ij} + \omega_{ij}$
  - thus, $\boldsymbol{Y}_i \sim N(\mathbb{X}_i\beta, \nu^2\mathbb{J}_i + \sigma^2\mathbb{H}_i + \tau^2\mathbb{I}_i)$
  - and, also, $\boldsymbol{\Omega}_i \sim N(\boldsymbol{0}, \nu^2\mathbb{J}_t + \sigma^2\mathbb{H}_t)$

- **Example 2** Assume a simple linear (regression) model with a random intercept and random slope (i.e., $\boldsymbol{z}_{ij} = (1, t_{ij})^\top$ and $\boldsymbol{w}_i \sim N_2(\boldsymbol{0}, \nu^2\mathbb{I})$), where $\mathbb{I} \in \mathbb{R}^{2\times 2}$ is a unit matrix and $\boldsymbol{w}_i = (w_{i1}, w_{i2})^\top$
  - observations $Y_i(t_{ij}) \equiv Y_{ij} = \mu(t_{ij}) + (w_{i1} + w_{i2}t_{ij}) + W_i(t)_{ij} + \omega_{ij}$
  - thus, $\boldsymbol{Y}_i \sim N(\mathbb{X}_i\beta, \nu^2\mathbb{M}_i + \sigma^2\mathbb{H}_i + \tau^2\mathbb{I}_i)$, where $\mathbb{M}_i = (1 + t_{ij}t_{ik})_{j,k=1}^{n_i}$
  - and, also, $\boldsymbol{\Omega}_i \sim N(\boldsymbol{0}, \nu^2\mathbb{M}_t + \sigma^2\mathbb{H}_t)$, where $\mathbb{M}_t = (1 + t_jt_k)_{j,k=1}^{n}$

Bayesian interpretation

- prior density for the random effects: $\boldsymbol{w}_i \sim g(\boldsymbol{w})$
- conditional density of the data: $\boldsymbol{Y}_i | \boldsymbol{w}_i \sim f(\boldsymbol{y}|\boldsymbol{w})$
- posterior density for the random effects

$$g(\boldsymbol{w}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{w})g(\boldsymbol{w})}{\int f(\boldsymbol{y}|\boldsymbol{w})g(\boldsymbol{w})\mathrm{d}\boldsymbol{w}}$$

- posterior mean of $g(\boldsymbol{w}|\boldsymbol{y})$ used as an estimate for $\boldsymbol{w}_i$
(still depends on the estimated parameters in $\widehat{\boldsymbol{\alpha}} \in \mathbb{R}^q$)