Topic 2: General Linear Models for Longitudinal Data

(I) $t_{ij} = t_j, i = 1,\cdots,n, j = 1,\cdots,m.$ ($N = nm$)

Let $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$, and $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ with $y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im} \end{pmatrix}, \varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{im} \end{pmatrix}$,

and $X_i = \begin{pmatrix} X_{i10} & X_{i11} & \cdots & X_{i1p} \\ X_{i20} & X_{i21} & \cdots & X_{i2p} \\ \vdots & \vdots & & \vdots \\ X_{im0} & X_{im1} & \cdots & X_{imp} \end{pmatrix}$ with $X_{ij0} = 1$.

The **general linear model (GLM)** is given by

$Y = X\beta + \varepsilon$, where $\varepsilon \sim (0, \sigma^2 V)$ with $V = diag(V_0,\cdots,V_0)$.

(a) Without making parametric assumptions about V_0.

(b) The parametric modeling approach of V_0.

(b1) The uniform correlation model: $V_0 = (1-\rho)I + \rho J$ with $J = \cancel{11^T}\ 1_m 1_m^T$.

Let $\varepsilon_i = u_i 1_m + Z_i, u_i\text{'s} \overset{iid}{\sim} (0, \tau_2^2) \perp\!\!\!\perp Z_i\text{'s} \overset{iid}{\sim} (0, \tau_1^2 I_m)$.

One can derive that $Var(\varepsilon_i) = \tau_1^2 I_m + \tau_2^2 1_m 1_m^T$

$$= (\tau_1^2 + \tau_2^2)((1-\rho)I_m + \rho J), \rho = \frac{\tau_2^2}{\tau_1^2 + \tau_2^2}.$$

(b2) The exponential correlation model: $V_{0jk} = e^{-\phi|t_j - t_k|} = \rho^{-|t_j - t_k|}$ with $\rho = e^{-\phi}$.

Example:

First order autoregressive model: (equally spaced time points)

$\varepsilon_{ij} = \rho \overset{\perp\!\!\!\perp}{\varepsilon_{ij-1} + Z_{ij}}$ with $Z_{ij}\text{'s} \overset{iid}{\sim} (0, (1-\rho^2)\sigma^2)$, $|\rho| < 1$.

$E[\varepsilon_{ij}] = 0$ & $Var(\varepsilon_{ij}) = \frac{(1-\rho^2)\sigma^2}{1-\rho^2} = \sigma^2$

$Cov(\varepsilon_{ij}, \varepsilon_{ij-k}) = \rho^k \sigma^2$

Generalization of the discrete-time first order autoregressive process:

$ARMV(p,q)$ process: $\varepsilon_{ij} = \sum_{r=1}^{p} \rho_r \varepsilon_{ij-r} + Z_{ij} + \sum_{s=1}^{q} \beta_s Z_{ij-s}$.

Further generalization: (This generalization can accommodate irregularly spaced time points t_{ij}'s .)

$\varepsilon_{ij} = \omega_i(t_j)$, where $\{\omega_i(t_j)\ ;\ j = 1,\cdots,m\}'s$ are realizations of mutually independent, continuous-time, and stationary process $\{\omega_i(t)\ , t \in R\}$ with a common covariance structure $r(u) = Cov(\omega_i(t), \omega_i(t-u))$.

Two-stage least-squares estimation for random effects models –

$Y_i = X_i\beta_i + \varepsilon_i$, where $\varepsilon_i \sim (0, \tau^2 I_m)$, $i = 1,\cdots,n$.

The first-stage analysis: $\hat{\beta}_i = (X_i^T X_i)^{-1} X_i^T Y_i$ with $\hat{\beta}_i = \beta_i + Z_i$ and $Z_i \sim (0, \tau^2 (X_i^T X_i)^{-1})$.

The second-stage analysis: $\beta_i = \beta + \delta_i$, where δ_i's $\overset{iid}{\sim} (0, \sigma^2 I_{p+1})$.

It implies that $\hat{\beta}_i = \beta + (Z_i + \delta_i)$ with $Z_i + \delta_i \sim (0, \sigma^2 I_{p+1} + \tau^2 (X_i^T X_i)^{-1})$ and,

hence, $\hat{\beta}$ can be obtained via fitting the second stage model.

Remark. Combine the first-stage and second-stage models into a single equation:

$Y_i = X_i\beta_i + \varepsilon_i = X_i\beta + \varepsilon_i^*$, where $\varepsilon_i^* = X_i\delta_i + \varepsilon_i \sim (0, \sigma^2 I_m + \tau^2 (X_i X_i^T))$.

Here, $\tilde{\beta}$ can be obtained via fitting the above model in a single-stage calculation.

Weighted Least-Squares Estimation – $Y = X\beta + \varepsilon$, $\varepsilon \sim (0, \sigma^2 V)$.

The weighted least squares estimator, say $\hat{\beta}_w$ of β is defined to be

$$\hat{\beta}_w = (X^T W X)^{-1} X^T W Y = \arg\min_{\beta} (Y - X\beta)^T W (Y - X\beta)$$

Remark.

(a) W^{-1} is called the working variance matrix.

(b) $E[\hat{\beta}_w] = \beta$ and $Var(\hat{\beta}_w) = \sigma^2[(X^TWX)^{-1}X^TW]V[WX(X^TWX)^{-1}]$.

(c) When V is nonsingular and $W = V^{-1}$, one can show that $\hat{\beta}_w$ is the best linear unbiased estimator (BLUE) of β.

Maximum likelihood Estimation under Gaussian assumptions:
The likelihood function of (β, σ^2, V) is

$L(\beta, \sigma^2, V) = (2\pi\sigma^2)^{\frac{-N}{2}} |V|^{\frac{-1}{2}} e^{\frac{-(Y-X\beta)^T V^{-1}(Y-X\beta)}{2\sigma^2}}$ with the log-likelihood function

$$l(\beta, \sigma^2, V) = \frac{-1}{2}(N\ln 2\pi + N\ln\sigma^2 + \ln|V| + \sigma^{-2}(Y - X\beta)^T V^{-1}(Y - X\beta)).$$

For given V, $\arg\max_{\beta} l(\beta, \sigma^2, V) = \hat{\beta}(V) = (X^TV^{-1}X)^{-1}(X^TV^{-1}Y)$.

Substituting $\hat{\beta}(V)$ into $l(\beta, \sigma^2, V)$,

$$\arg\max_{\sigma^2} l(\hat{\beta}(V), \sigma^2, V) = \hat{\sigma}^2(V) = \frac{1}{N}(Y - X\hat{\beta}(V))^T V^{-1}(Y - X\hat{\beta}(V)).$$

Again, substituting $\hat{\beta}(V)$ and $\hat{\sigma}^2(V)$ into $l(\beta, \sigma^2, V)$, the log-likelihood function is

reduced to $l(\hat{\beta}(V), \hat{\sigma}^2(V), V) = \frac{-1}{2}(N(\ln 2\pi + 1) + N\ln\hat{\sigma}^2(V) + \ln|V|)$.

One has $\hat{V} = \arg\max_{V} l(\hat{\beta}(V), \hat{\sigma}^2(V), V)$.

The MLE of (β, σ^2, V) is then defined to be $(\hat{\beta}(\hat{V}), \hat{\sigma}^2(\hat{V}), \hat{V}) = \arg\max_{(\beta, \sigma^2, V)} l(\beta, \sigma^2, V)$.

Remark.
1. A sensible strategy to combat inconsistent estimators for (σ^2, V), an over-elaborate model for the mean response profile is used in estimating the covariance structure of the data.
2. An unbiased estimator for $\hat{\sigma}^2(V)$ requires a divisor $(N - (p+1))$ rather than N.

Restricted Maximum likelihood Estimation – A way for estimating variance components in a GLM $Y = X\beta + \varepsilon$, where $\varepsilon \sim (0, \sigma^2 V)$.

Remark. The restricted maximum likelihood (REML) estimator is defined as a MLE based on a linearly transformed set of data $Z = B^TY$ so that the distribution of Z does not depend on β.

Let $A = I - X(X^T X)^{-1} X^T$ and B be the $N \times (N-(p+1))$ matrix such that $BB^T = A$ and $B^T B = I_{N-(p+1)}$.

For any fixed V, the MLE of β is $\hat{\beta}(V) = GY$ with $G = (X^T V^{-1} X)^{-1} X^T V^{-1}$.

Facts:

$\hat{\beta}(V) \sim N_{p+1}(\beta, \sigma^2 (X^T V^{-1} X)^{-1})$, $Z \sim N_{N-(p+1)}(0, \sigma^2 (B^T V B))$, and $\hat{\beta}(V) \perp Z$. (Using $B^T X = B^T B B^T X = B^T A X = B^T 0 = 0$.)

Let $U = \begin{pmatrix} Z \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} B^T \\ G \end{pmatrix} Y$.

From $f_U(u) = f_Y(y) \left| \begin{pmatrix} B^T \\ G \end{pmatrix} \right|^{-1}$, where $\left| \begin{pmatrix} B^T \\ G \end{pmatrix} \right| = | GG^T - GBB^T G^T |^{\frac{1}{2}} = | X^T X |^{\frac{-1}{2}}$, one can derive that

$$f_Z(z) = (2\pi\sigma^2)^{\frac{-(N-(p+1))}{2}} \exp\left(\frac{-1}{2\sigma^2}(y - X\hat{\beta}(V))^T V^{-1}(y - X\hat{\beta}(V))\right) | X^T X |^{\frac{1}{2}} | V |^{\frac{-1}{2}}$$

$\cdot | X^T V^{-1} X |^{\frac{1}{2}}$, and the log restricted likelihood function:

$$l(\hat{\beta}(V), \sigma^2, V) = \frac{-(N-(p+1))(\ln(2\pi\sigma^2))}{2} + \frac{1}{2}(\ln(| X^T X |) - \ln(| V |) + \ln(| X^T V^{-1} X |))$$

$$-\frac{1}{2\sigma^2}(y - X\hat{\beta}(V))^T V^{-1}(y - X\hat{\beta}(V))).$$

Then, the RMLE of σ^2 is derived to be

$$\hat{\sigma}^2(V) = \arg\max_{\sigma^2} l(\hat{\beta}(V), \sigma^2, V) = (Y - X\hat{\beta}(V_0))^T V^{-1}(Y - X\hat{\beta}(V))/(N-(p+1))$$

By substituting $\hat{\sigma}^2(V)$ into $l(\hat{\beta}(V), \sigma^2, V)$, one has $l(\hat{\beta}(V), \hat{\sigma}^2(V), V) =$

$$\frac{-(N-(p+1))}{2}(\ln(2\pi\hat{\sigma}^2(V)) + 1) + \frac{1}{2}(\ln(| X^T X |) - \ln(| V |) + \ln(| X^T V^{-1} X |)).$$

The estimator, say, $\hat{V}$, for V can be obtained via maximizing $l(\hat{\beta}(V), \hat{\sigma}^2(V), V)$.

Finally, substituting $\hat{V}$ into $\hat{\beta}(V)$ and $\hat{\sigma}^2(V)$, the REML estimators for (β, σ^2, V) are defined to be $(\hat{\beta}(\hat{V}), \hat{\sigma}^2(\hat{V}), \hat{V})$.

Robust Estimation of standard errors –

The robust approach to inference of β is to use the generalized least-squares estimator $\hat{\beta}_w = (X^T W X)^{-1} X^T W Y$ in conjunction with an estimated variance-covariance matrix $\hat{R}_w = (X^T W X)^{-1} X^T W \hat{V}^* W X (X^T W X)^{-1}$, where $\hat{V}^*$ is consistent for $\sigma^2 V$ ~~whatever the true covariance structure.~~

Remark.

(a) W^{-1} is called the working variance matrix, which is used to distinguish from the true variance matrix.

(b) A poor choice of W will affect only the efficiency of inferences for β not their validity.

(c) $\hat{\beta}_w \approx N_{p+1}(\beta, \hat{R}_w)$.

(Ⅱ) t_{ij}'s are not occurring at a common set of times:

Let $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$, $X = \begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$, and $t = \begin{pmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_n^T \end{pmatrix}$ with $Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{im_i} \end{pmatrix}$, $\varepsilon_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{im_i} \end{pmatrix}$, $t_i = \begin{pmatrix} t_{i1} \\ t_{i2} \\ \vdots \\ t_{im_i} \end{pmatrix}$,

and $X_i = \begin{pmatrix} X_{i10} & X_{i11} & \cdots & X_{i1p} \\ X_{i20} & X_{i21} & \cdots & X_{i2p} \\ \vdots & \vdots & & \vdots \\ X_{im_i0} & X_{im_i1} & \cdots & X_{im_ip} \end{pmatrix}$, with $\mathrm{X}_{\mathrm{ij0}} = 1, i = 1, \cdots, n, j = 1, \cdots, m_i$.

$(N = \sum_{i=1}^{n} m_i)$

Parametric models are used to model the covariance structure of Y_i ~~with~~ , where Y_i's ~~being~~ are mutually independent with $(X_i^T\beta, V_i(t_i,\alpha))$ ~~and~~ ($\dim(\alpha)=q$), i.e. $Y = X\beta + \varepsilon$ ~~, where~~ with $\varepsilon \sim (0, V(t,\alpha))$ and $V(t,\alpha) = diag(V_1(t_1,\alpha),\cdots,V_n(t_n,\alpha))$.

Remark.
(a) ε_i's are assumed to be sampled from independent copies of an underlying continuous-time stochastic process $\{\varepsilon(t), t \in R\}$.
(b) The properties of each model for the stochastic process $\{\varepsilon(t)\}$ can be described by the covariance function or variogrm $r(u) = \frac{1}{2}E[(\varepsilon(t)-\varepsilon(t-u))^2]$ when $\varepsilon(t)$ is stationary.

The possible sources of random variation in longitudinal data –

(a) Random effects: Individual's behavior may show stochastic variation between units.
(b) Serial correlation: Any unit's observed measurement profile may be a response to time-varying stochastic process operating within that unit.
(c) Measurement error: The measurement process may itself add a component of variation to the data.

Consider $Y = X\beta + \varepsilon$ with $\varepsilon_i = D_i U_i + \omega_i + z_i, i = 1,\cdots,n$, where

(a1) $D_i = \begin{pmatrix} d_{i1}^T \\ \vdots \\ d_{im_i}^T \end{pmatrix}$ with d_{ij}'s being r-element vectors of explanatory variables attached to individual measurements, and $U_1,\cdots,U_n$ are mutually independent random effects with $(0, G_{r\times r})$.

(b1) $\omega_i = \begin{pmatrix} \omega_i(t_{i1}) \\ \omega_i(t_{i2}) \\ \vdots \\ \omega_i(t_{im_i}) \end{pmatrix}$ with ω_i's being independently sampled from a stationary process with mean 0, variance σ^2, and correlation function $\rho(t)$, i.e., ω_i's are independently from a population with $(0, \sigma^2 H_i)$ with the serial correlation $h_{jk} = \rho(|t_{ij} - t_{ik}|)$.

(c1) z_i's are independent measurement errors with $(0, \tau^2 I_{m_i \times m_i})$.

diag$(D_1 G D_1^T, \cdots, D_n G D_n^T)$

It implies from the above model that $\varepsilon \sim (0, \cancel{DGD^T} + \sigma^2 H + \tau^2 I_N)$, where $\cancel{D = \begin{pmatrix} D_1 \\ \vdots \\ D_n \end{pmatrix}}$,

and $H = diag(H_1, \cdots, H_n)$.

Case1. Pure serial correlation: $\varepsilon \sim (0, \sigma^2 H)$.
The corresponding variogram is $r(u) = \sigma^2(1 - \rho(u))$.

Remark. Typically, $\rho(u)$ decreases with increasing time-separation u, which implies that $r(0) = 0$ and $r(u) \to \sigma^2$ as $u \to \infty$. Small $r(u)$ implies smoother error process $\varepsilon(t)$.

Example:
(a1) Exponential correlation model: $\rho(u) = \exp(-\phi u)$ with $\phi > 0$, which is continuous but not differentiable at 0.
(a2) Gaussian correlation model: $\rho(u) = \exp(-\phi u^2)$, $\phi > 0$.

Ante-dependence model of order p or the pth order Markov model:

The conditional distribution of ε_{ij}, given $\varepsilon_{ij-1}, \cdots, \varepsilon_{i1}$, depends only on $\varepsilon_{ij-1}, \cdots, \varepsilon_{ij-p}$, i.e.

$$f(\varepsilon_{i1}, \cdots, \varepsilon_{im_i}) = f_0(\varepsilon_{i1}, \cdots, \varepsilon_{ip}; \alpha) \cdot \left(\prod_{j=p+1}^{m_i} f_c(\varepsilon_{ij} \mid \varepsilon_{ij-1}, \cdots, \varepsilon_{ij-p}; \alpha)\right).$$

Remark.
(a) It may not easy to deduce $f_0(\cdot)$. In practical implementation, using the product of the $(m_i - p)$'s conditional densities $f_c(\cdot)$ to estimate α.
(b) When n is large and p is small, the loss of information is small.
(c) Ante-dependence models are more appealing for equally spaced data, less so for un-equally spaced data. In addition, they do not cope easily with data for which the times of measurements are not common to all units.

$\boxed{\text{EXCEPTION}}$: $\varepsilon_i \sim N_{m_i}(0, \sigma^2 H_i)$, $h_{ijk} = \exp(-\phi \mid t_{ij} - t_{ik} \mid)$ with $\phi > 0$.

The joint p.d.f. of $(\varepsilon_{i1}, \cdots, \varepsilon_{im_i})$ can be derived by using the conditional p.d.f. of

$\varepsilon_{ij} \mid \varepsilon_{ij-1} \overset{d}{\sim} N(\alpha_j \varepsilon_{ij-1}, \sigma^2(1-\alpha_j^2))$, where $\alpha_j = \exp(-\phi \mid t_{ij} - t_{ij-1} \mid)$ and $\varepsilon_{i1} \sim N(0, \sigma^2)$.

Case2. Serial correlation plus measurement error: $\varepsilon \sim (0, \sigma^2 H + \tau^2 I_N)$.

The corresponding variogram is $r(u) = \tau^2 + \sigma^2(1-\rho(u))$ for $u \geq 0$ with $r(0) = \tau^2 > 0$.

Case3. Random intercept plus serial correlation plus measurement error:

$\varepsilon \sim (0, \upsilon^2 J + \sigma^2 H + \tau^2 I_N)$. The corresponding variogram is $r(u) = \tau^2 + \sigma^2(1-\rho(u))$

for $u \geq 0$, and $Var(\varepsilon_{ij}) = \upsilon^2 + \sigma^2 + \tau^2 > r(u)$ as $u \to \infty$.

$\mathrm{diag}(D_1 G D_1^T, \ldots, D_n G D_n^T)$

Case4. Random effects plus measurement error: $\varepsilon \sim (0, \cancel{DGD^T} + \tau^2 I_N)$.

$$\rho(\varepsilon_{ij}, \varepsilon_{ik}) = \frac{d_{ij}^T G d_{ik}}{[(d_{ij}^T G d_{ij} + \tau^2)(d_{ik}^T G d_{ik} + \tau^2)]^{1/2}}.$$

Remark: Random intercept plus measurement error, $\varepsilon \sim (0, \upsilon^2 J + \tau^2 I)$, implies that $\rho(u) = \dfrac{\upsilon^2}{\upsilon^2 + \tau^2}$, which is a uniform correlation structure and is sometimes called "a split-plot model", where subject is treated as a sub-plot.

Model fitting process –

(a) Formulation: Choosing the general form of the model. (continuation of the EDA)
(b) Estimation : Attaching numerical values to parameters.
(c) Inference : Calculating confidence intervals or testing hypotheses about the interesting parameters.
(d) Diagnosis: Checking the considered model.

Maximum likelihood estimation under Gaussian assumption: $Y \sim N_N(X\beta, \sigma^2 V(\alpha))$

For given α, the estimator $\hat{\beta}(\alpha)$ of β is derived as $\hat{\beta}(\alpha) = (X^T V^{-1}(\alpha) X)^{-1} X^T V^{-1}(\alpha) Y$.

estimator

The REML of σ^2 is $\hat{\sigma}^2(\alpha) = (Y - X\hat{\beta}(\alpha))^T V^{-1}(\alpha)(Y - X\hat{\beta}(\alpha))/(N-(p+1))$, and the REML estimator of α can be obtained via maximizing

$$
l^*(\alpha) = \cancel{\frac{-1}{2}\{(N-(p+1))\ln\hat{\sigma}^2(\alpha)+\ln|V(\alpha)|+\ln(|X^T V^{-1}(\alpha)X|)\}}
$$

$$
= \frac{-1}{2}\{(N-(p+1))\ln\hat{\sigma}^2(\alpha)+\sum_{i=1}^{n}\ln|V_i(t_i,\alpha)|+\sum_{i=1}^{n}\ln(|X_i^T V_i^{-1}(\alpha)X_i|)\}
$$

The resulting REML estimators of β and σ^2 are $\hat{\beta}(\hat{\alpha})$ and $\hat{\sigma}^2(\hat{\alpha})$, respectively.

Facts: $\hat{\beta}(\alpha) \sim N_p(\beta, \sigma^2 (X^T V^{-1}(\alpha) X)^{-1})$ and $\hat{\beta}(\alpha) \approx N_p(\beta, \hat{\sigma}^2(\hat{\alpha})\ (X^T V^{-1}(\hat{\alpha}) X)^{-1})$.

Confidence region of $\phi = D\beta$:

An approximated $(1-\alpha)$ confidence region for $\phi = D\beta$, $rank(D) = r \le (p+1)$, can be constructed as: $\{\phi : (\phi-\hat{\phi})^T\ D(X^T V^{-1}(\hat{\alpha})X)\ D^T(\phi-\hat{\phi}) / \hat{\sigma}^2(\hat{\alpha}) \le \chi^2_{r,1-\alpha}\}$.

Hypothesis test: $\begin{cases} H_0 : \phi = \phi_0 \\ H_A : \phi \ne \phi_0 \end{cases}$

(a) An approximated testing procedure is based on the test statistic:

$\chi^2 = (\phi-\phi_0)^T\ D(X^T V^{-1}(\hat{\alpha})X)\ D^T(\phi-\phi_0)/\hat{\sigma}^2(\hat{\alpha})$.

(b) An alternative testing procedure is based on the test statistic:

$-2\ln\left(\dfrac{\max_{\{\beta,\sigma^2,\alpha:\phi=\phi_0\}} L(\beta,\sigma^2,\alpha)}{\max_{\{\beta,\sigma^2,\alpha\}} L(\beta,\sigma^2,\alpha)}\right)$, which is approximately distributed as χ^2_{p+1-r} under H_0.

Estimation of individual trajectories –

$Y_i = X_i\beta + R_i + z_i$, $i = 1,\cdots,n$, where $R_i = U_i^* + \omega_i$ with $U_i^* = D_i U_i$.

where U_i's $\overset{iid}{\sim} N_r(0_{r\times 1}, \upsilon^2 G_i), \omega_i$'s $\overset{independent}{\sim} N_{m_i}(0, \sigma^2 H_i)$, and z_i's $\overset{independent}{\sim} N_{m_i}(0, \tau^2 I_{m_i})$.

Predict $Y_i(t)$ via $\hat{Y}_i(t) = X_i\hat{\beta} + \hat{R}_i(t)$ with $\hat{R}_i(t) \triangleq \hat{U}_i^*(t) + \hat{\omega}_i(t)$.

Using $\begin{pmatrix} R_i \\ Y_i \end{pmatrix} \sim N_{2m_i}\left(\begin{pmatrix} 0 \\ X_i^T\beta \end{pmatrix}, \begin{pmatrix} \upsilon^2 D_i G_i D_i^T + \sigma^2 H_i & \upsilon^2 D_i G_i D_i^T + \sigma^2 H_i \\ \upsilon^2 D_i G_i D_i^T + \sigma^2 H_i & \tau^2 \mathrm{I}_{m_i} + \upsilon^2 G_i + \sigma^2 H_i \end{pmatrix}\right)$, and

$R_i | Y_i = y_i \sim N_{m_i}(\mu_{R_i}, \Sigma_{R_i})$,

where $\mu_{R_i} = (\upsilon^2 D_i G_i D_i^T + \sigma^2 H_i)(\tau^2 I_{m_i} + \upsilon^2 D_i G_i D_i^T + \sigma^2 H_i)^{-1}(y_i - X_i^T \beta)$, and

$\Sigma_{R_i} = (\upsilon^2 D_i G_i D_i^T + \sigma^2 H_i)(I_{m_i} - (\tau^2 I_{m_i} + \upsilon^2 D_i G_i D_i^T + \sigma^2 H_i)^{-1}(\upsilon^2 D_i G_i D_i^T + \sigma^2 H_i))$.

It implies that R_i can be estimated by

$\hat{\mu}_{R_i} = (\hat{\upsilon}^2 D_i \hat{G}_i D_i^T + \hat{\sigma}^2 \hat{H}_i)\,(\hat{\tau}^2 I_i + \hat{\upsilon}^2 D_i \hat{G}_i D_i^T + \hat{\sigma}^2 \hat{H}_i)^{-1}\,(y_i - X_i^T \hat{\beta})$.

Repeated measures:

$y_{hij} = \beta_h + r_{hj} + U_{hi} + z_{hij}$, $h = 1, \cdots, g$, $i = 1, \cdots, n_h$, $j = 1, \cdots, m$,

where

(1) β_h is the main effect for the hth treatment and r_{hj} is the interaction effect between the hth treatment and the jth time with $\sum_{j=1}^{m} r_{hj} = 0 \ \forall h$.

(2) $U_{hi} \overset{iid}{\sim} N(0, \upsilon^2)$ and $z_{hij} \overset{iid}{\sim} N(0, \sigma^2)$ are independent.

It implies that $Y_{hi} = (y_{hi1}, \cdots, y_{him})^T \sim N_m(\beta_h 1_m + r_h, \sigma^2 I_m + \upsilon^2 J)$ with $r_h = (r_{h1}, \cdots, r_{hm})^T$.

Remark.

(a) The derived approach breaks down when the array, y_{hij}, is incomplete because of missing values, or when the measurement times are not common.

(b) The rationale behind the analysis is to regard time as a factor with m levels in hierarchical design with units as sub-plots. However, the allocation of times to the m observations within each unit is not randomized.

ANOVA table:

Source of variation	Sum of squares	Degrees of freedom
Between treatments	$BTSS_1 = \sum_{h=1}^{g} (n_h m)(\bar{y}_{h\cdot\cdot} - \bar{y}_{\cdots})^2$	$(g-1)$
Whole plot residuals	$RSS_1 = TSS_1 - BTSS_1$	$(\sum_{h=1}^{g} n_h - g)$

Whole plot total	$\underline{TSS_1} = \sum_{h=1}^{g}\sum_{i=1}^{n_h} m(\bar{y}_{hi\cdot} - \bar{y}_{\dots})^2$	$(\sum_{h=1}^{g} n_h - 1)$
Between times	$BTSS_2 = (\sum_{h=1}^{g} n_h)\sum_{j=1}^{n}(\bar{y}_{\cdot j\cdot} - \bar{y}_{\dots})^2$	$m-1$
Treatment by time interactions	$ISS_2 = \sum_{h=1}^{g}\sum_{j=1}^{m} n_h(\bar{y}_{h\cdot j} - \bar{y}_{h\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\dots})^2$	$(m-1)\times(g-1)$
Split-plot residual	$RSS_2 = TSS_2 - \underline{BTSS_2 - ISS_2 - TSS_1}$	$(m\sum_{h=1}^{g} n_h - g)\times(m-1)$
Split-plot total	$\underline{TSS_2} = \sum_{h=1}^{g}\sum_{i=1}^{n_h}\sum_{j=1}^{m}(y_{hij} - \bar{y}_{\dots})^2$	$m(\sum_{h=1}^{g} n_h) - 1$

Hypotheses: $\begin{cases} H_0 : \beta_1 = \cdots = \beta_g & F = \dfrac{MBTSS_1}{MRSS_1} \\ H_A : r_{1j} = \cdots = r_{gj} \ , j = 1,\cdots,m. & F = \dfrac{MISS_2}{MRSS_2} \end{cases}$

Analysis of variance methods–

$y_{hij} = \mu_{hj} + \varepsilon_{hij}, h = 1,\cdots,g, \mathrm{i}=1,\cdots,\mathrm{n_h}, \mathrm{j} = 1,\cdots,m,$ where μ_{hj} is the mean response profile at time t_j in the hth group.

Let $\mu_h = (\mu_{h1},\cdots,\mu_{hm})^T$.

Time-by-time ANOVA: A time-by-time ANOVA consists of m separate analyses, one for each subset of data corresponding to each time t_j .

ANOVA table for the jth time point:

Source of variation	Sum of squares	d.f.	F-statistic
Between treatments	$SSB_j = \sum_{h=1}^{g} n_g(\bar{y}_{h\cdot j} - \bar{y}_{\cdot\cdot j})^2$	$(g-1)$	$F = \dfrac{MSB_j}{MSE_j}$
Residuals	$SSE_j = \sum_{h=1}^{g}\sum_{i=1}^{n_h}(\bar{y}_{hij} - \bar{y}_{h\cdot j})^2$	$(\sum_{h=1}^{g} n_h - g)$	
Total	$SST_j = \sum_{h=1}^{g}\sum_{i=1}^{n_h}(\bar{y}_{hij} - \bar{y}_{\cdot\cdot j})^2$	$m(\sum_{h=1}^{g} n_h) - 1$	

Drawbacks:

(a) It cannot address questions concerning the treatment effects which relate to the longitudinal development of the mean response profiles.

(b) The inferences made within each of the m separate analyses are not independent, nor is it clear how they should be combined.

Remark. One possible solution is to use the derived variables of $y_{hi} = (y_{hi1}, \cdots, y_{him})^T$, $h = 1, \cdots, g$; $i = 1, \cdots, n_h$, i.e.~~,~~ a scalar-value function, say, $u_{hi} = u(y_{hi})$.