Chapter 1

Linear Mixed Effects Models

1.1 The General Linear Model

1.1.1 The Aitken Model and GLS

– When Y is continuous random variable, it is often assumed that the random variable Y follows the normal distribution $Y \sim N(\mu, \sigma^2)$ with the structure of the expected value $\mu(\mathbf{x}, \boldsymbol{\beta})$ being *linear*

$$\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = \mathbf{x}'\boldsymbol{\beta}. \tag{1.1}$$

– Considered statistical model is usually assumed to hold for every unit of the population Ω in question. For the unit $i \in \Omega$, the linear model under normality can be defined by the equations

$$Y_i \sim N(\mu_i, \sigma^2), \tag{1.2a}$$

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} = \mathbf{x}_i'\boldsymbol{\beta}, \tag{1.2b}$$

where $\mathbf{x} = (x_{i1}, x_{i2}, \ldots, x_{ip})'$ are observed or set values of the explanatory variables for the unit i, and Y_i is an observable random variable associated to the unit i.

– It is useful to define the considered statistical model to the selected sampling units $i = 1, 2, \ldots, n$. For example for the units $i = 1, 2, \ldots, n$, the linear model under normality with independence is defined by equations

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I}), \tag{1.3a}$$
$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \tag{1.3b}$$

where

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \text{ is } n \times 1 \text{ observable random vector,}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \text{ is } n \times 1 \text{ unknown vector of expected values,}$$

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 1 \end{pmatrix} \text{ is } n \times n \text{ identity matrix,}$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} = \left(\mathbf{1} : \mathbf{x}_{(1)} : \mathbf{x}_{(2)} : \cdots : \mathbf{x}_{(p)}\right) = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ & & & \vdots & \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix}$$

is $n \times (p+1)$ model matrix containing the values of the explanatory variables.

Note that in these notes (and in statistics generally!) the vector $\mathbf{y}$ has double meaning. It is used for denoting the random variable $\mathbf{y} = (Y_1, Y_2, \dots, Y_n)'$ but also as the vector of realizations of the random variables: $\mathbf{y} = (y_1, y_2, \dots, y_n)'$.

– A linear normal model is often expressed in a form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \tag{1.4}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ is called as unknown vector of random errors with the expected value $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0} = (0, 0, \dots, 0)'$ and the covariance matrix $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$.

– If the elements of the random vector $\mathbf{y}$ are assumed to be correlated with each others, then the linear normal model is defined by equations

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2\mathbf{V}), \tag{1.5a}$$

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}, \tag{1.5b}$$

or as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{V}), \tag{1.6}$$

where the covariance matrix $\mathrm{Cov}(\mathbf{y}) = \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$ has in general the form

$$\sigma^2\mathbf{V} = \sigma^2 \begin{pmatrix} 1 & \varrho_{12} & \dots & \varrho_{1n} \\ \varrho_{21} & 1 & \dots & \varrho_{2n} \\ & & \vdots & \\ \varrho_{n1} & \varrho_{n2} & \dots & 1 \end{pmatrix}, \quad \text{e.g., the correlation } \mathrm{Cor}(Y_1, Y_2) = \frac{\varrho_{12}}{\sigma^2}.$$

– The model can be called as the Aitken model.

– The Aitken model can be transformed to the ordinal linear model by linear transformation

$$\begin{aligned} \mathbf{V}^{-\frac{1}{2}}\mathbf{y} &= \mathbf{V}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-\frac{1}{2}}\boldsymbol{\varepsilon} \\ \mathbf{y}_* &= \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*, \end{aligned} \tag{1.7}$$

where $\mathbf{V}^{-1} = \mathbf{V}^{-\frac{1}{2}}\mathbf{V}^{-\frac{1}{2}}$, and hence

$$\mathbf{y}_* \sim N(\boldsymbol{\mu}_*, \sigma^2\mathbf{I}), \tag{1.8a}$$

$$\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}. \tag{1.8b}$$

– In the Aitken model, estimation of $\boldsymbol{\beta}$ is based on the generalized least squares

$$\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \tag{1.9}$$

– Based on the generalized least squares and, more general, theory of linear models, we have

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{1.10}$$

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{1.11}$$

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - \operatorname{rank}(\mathbf{X})} \\ &= \frac{\mathbf{y}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}}{n - \operatorname{rank}(\mathbf{X})} = \frac{\mathbf{y}'\dot{\mathbf{M}}\mathbf{y}}{n - \operatorname{rank}(\mathbf{X})},\end{aligned} \tag{1.12}$$

where

$$\dot{\mathbf{M}} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}.$$

– In the Aitken model, we can consider general linear hypotheses

$$H_0 : \mathbf{K}'\boldsymbol{\beta} = \mathbf{0},$$
$$H_1 : \mathbf{K}'\boldsymbol{\beta} \neq \mathbf{0},$$

where $\mathbf{K}' \in \mathbb{R}^{q,(p+1)}$.

– If H_0 holds,

$$\mathbf{K}'\hat{\boldsymbol{\beta}} \sim N(\mathbf{0}, \sigma^2\mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K}) \tag{1.13}$$

$$Q = \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'(\mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{\sigma^2} \sim \chi^2_q, \tag{1.14}$$

where χ^2_q denotes the central χ^2-distribution with degrees freedom q.

– If H_0 holds, Wald statistic

$$\begin{aligned} W = \frac{Q/q}{\frac{\mathbf{y}'\dot{\mathbf{M}}\mathbf{y}}{\sigma^2}/n-(p+1)} &= \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'(\tilde{\sigma}^2\mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{q} \\ &= \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}})'\left(\widetilde{\mathrm{Cov}}(\mathbf{K}'\hat{\boldsymbol{\beta}})\right)^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}}}{q} \sim F_{q,n-(p+1)}, \end{aligned} \tag{1.15}$$

where $F_{q,n-(p+1)}$ denotes the F-distribution with degrees freedoms q and $n-(p+1)$.

– In practice, the covariance matrix $\mathbf{V}$ may contain some unknown parameters $\boldsymbol{\theta}$, denoted as $\mathbf{V}_{\boldsymbol{\theta}}$, which need to be estimated by the (restricted) maximum likelihood method.

– Under the covariance matrix $\mathbf{V}_{\boldsymbol{\theta}}$, hypotheses testing can also be based on likelihood ratio type of statistic.

1.1.2 The Best Linear Unbiased Prediction

– Consider prediction of new observation

$$y_f = \mathbf{x}_f'\boldsymbol{\beta} + \varepsilon_f.$$

– In the Aitken's general linear model, we assume that

$$\mathrm{E}\begin{pmatrix}\boldsymbol{\varepsilon}\\ \varepsilon_f\end{pmatrix} = \begin{pmatrix}\mathbf{0}\\ 0\end{pmatrix}, \qquad \mathrm{Cov}\begin{pmatrix}\boldsymbol{\varepsilon}\\ \varepsilon_f\end{pmatrix} = \sigma^2\begin{pmatrix}\mathbf{V} & \mathbf{w}\\ \mathbf{w}' & v_f\end{pmatrix},$$

– Hence

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ y_f\end{pmatrix} = \begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{x}_f'\boldsymbol{\beta}\end{pmatrix}, \qquad \mathrm{Cov}\begin{pmatrix}\mathbf{y}\\ y_f\end{pmatrix} = \sigma^2\begin{pmatrix}\mathbf{V} & \mathbf{w}\\ \mathbf{w}' & v_f\end{pmatrix}.$$

– Under the assumption of normality, we have

$$\begin{pmatrix}\mathbf{y}\\ y_f\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{x}_f'\boldsymbol{\beta}\end{pmatrix}, \sigma^2\begin{pmatrix}\mathbf{V} & \mathbf{w}\\ \mathbf{w}' & v_f\end{pmatrix}\right).$$

– A linear predictor $\mathbf{g}'\mathbf{y}$ of y_f is said to be *linear unbiased predictor* if (and only if)

$$\mathrm{E}(\mathbf{g}'\mathbf{y}) = \mathbf{x}_f'\boldsymbol{\beta} \qquad \text{for all } \boldsymbol{\beta} \in \mathbb{R}_{p+1}, \tag{1.16}$$

i.e., the expected prediction error is zero.

– It is easy to confirm that (1.16) is equivalent to $\mathbf{g}'\mathbf{X} = \mathbf{x}'_f$.

– A linear unbiased predictor $\mathbf{g}'\mathbf{y}$ is said to the *best linear unbiased predictor*, BLUP, of y_f, if the inequality $\mathrm{Var}(y_f - \mathbf{g}'\mathbf{y}) \leq \mathrm{Var}(y_f - \mathbf{f}'\mathbf{y})$ holds for every linear unbiased predictor $\mathbf{f}'\mathbf{y}$ of y_f.

– The fundamental equations of the BLUP:
The linear predictor $\mathbf{g}'\mathbf{y}$ is the best linear unbiased predictor of y_f if and only if the vector $\mathbf{g}'$ satisfies the equations

$$\mathbf{g}'(\mathbf{X} : \mathbf{VM}) = (\mathbf{x}'_f : \mathbf{w}'\mathbf{M}), \tag{1.17}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

– The BLUP of y_f has the forms

$$\begin{aligned}\hat{y}_f &= \mathbf{x}'_f\hat{\boldsymbol{\beta}} + \mathbf{w}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{x}'_f(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + \mathbf{w}'\mathbf{V}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}\end{aligned} \tag{1.18}$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

– For the BLUP $\hat{y}_f$, the variance of the prediction error is

$$\begin{aligned}\mathrm{Var}(e_f) &= \mathrm{Var}(y_f - \hat{y}_f)\\ &= \sigma^2\left(v_f - \mathbf{w}'\mathbf{V}^{-1}\mathbf{w} + (\mathbf{x}'_f - \mathbf{w}'\mathbf{V}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{x}_f - \mathbf{X}'\mathbf{V}^{-1}\mathbf{w})\right) \qquad (1.19)\\ &= \sigma^2\Sigma_{e_f}.\end{aligned}$$

– Thus under normality

$$y_f - \hat{y}_f \sim N(0, \sigma^2\Sigma_{e_f}),$$

and hence

$$Z = \frac{y_f - \hat{y}_f}{\sqrt{\sigma^2\Sigma_{e_f}}} \sim N(0,1). \qquad (1.20)$$

– Also asymptotically

$$Z = \frac{y_f - \hat{y}_f}{\sqrt{\tilde{\sigma}^2\Sigma_{e_f}}} \sim N(0,1). \qquad (1.21)$$

– Thus for $z_{\alpha/2}$ with $P(Z \geq z_{\alpha/2}) = \alpha/2$, we have asymptotically

$$\begin{aligned}
P\left(|Z| \leq z_{\alpha/2}\right) &= 1-\alpha \\
P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) &= 1-\alpha \\
P\left(-z_{\alpha/2} \leq \frac{(y_f - \hat{y}_f)}{\sqrt{\tilde{\sigma}^2 \Sigma_{e_f}}} \leq z_{\alpha/2}\right) &= 1-\alpha \qquad (1.22) \\
P\left(-z_{\alpha/2}\sqrt{\tilde{\sigma}^2 \Sigma_{e_f}} \leq y_f - \hat{y}_f \leq z_{\alpha/2}\sqrt{\tilde{\sigma}^2 \Sigma_{e_f}}\right) &= 1-\alpha \\
P\left(\hat{y}_f - z_{\alpha/2}\sigma\sqrt{\tilde{\sigma}^2 \Sigma_{e_f}} \leq y_f \leq \hat{y}_f + z_{\alpha/2}\sqrt{\tilde{\sigma}^2 \Sigma_{e_f}}\right) &= 1-\alpha.
\end{aligned}$$

– Thus the asymptotic $100(1-\alpha)\%$ prediction interval for the new observation y_f is

$$\left[\hat{y}_f - z_{\alpha/2}\sqrt{\tilde{\sigma}^2 \Sigma_{e_f}}, \hat{y}_f + z_{\alpha/2}\sqrt{\tilde{\sigma}^2 \Sigma_{e_f}}\right]. \qquad (1.23)$$

– If the covariance matrix $\begin{pmatrix} \mathbf{V} & \mathbf{w} \\ \mathbf{w}' & v_f \end{pmatrix}$ depends on unknown parameters $\boldsymbol{\theta}$, then in above everything should be replace by the estimates of the quantities $\mathbf{V}, \mathbf{w}, v_f$.

1.1.3 Time Series Regression Models

– Let us consider the linear model

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_p x_{tp} + \varepsilon_t, \tag{1.24}$$

where index t denotes time.

– Often in time series, the random error terms ε_t are correlated with past and future observations, i.e., $Cov(\varepsilon_{t-s}, \varepsilon_t) = Cov(\varepsilon_t, \varepsilon_{t+s}) = \sigma^2 \rho_s$.

– For example, ε_t may follow first order auto-regressive AR(1) process

$$\varepsilon_t = \phi \varepsilon_{t-1} + u_t, \tag{1.25}$$

with white noise u_t being independent normally distributed $u_t \sim N(0, \sigma^2)$.

– Under AR(1) process, the matrices $\mathbf{V}$ and $\mathbf{w}$ are

$$\mathbf{V} = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \ldots & \phi^{t-2} & \phi^{t-1} \\ \phi & 1 & \phi & \ldots & \phi^{t-3} & \phi^{t-2} \\ & \vdots & & & & \\ \phi^{t-2} & \phi^{t-3} & \phi^{t-4} & \ldots & 1 & \phi \\ \phi^{t-1} & \phi^{t-2} & \phi^{t-3} & \ldots & \phi & 1 \end{pmatrix}, \qquad \mathbf{w} = \frac{\sigma^2}{1-\phi^2} \begin{pmatrix} \phi^t \\ \phi^{t-1} \\ \vdots \\ \phi^2 \\ \phi \end{pmatrix}. \tag{1.26}$$

– More generally, we may assume that the random error term ε_t follows ARMA(p,q) process

$$\varepsilon_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \cdots + \phi_p\varepsilon_{t-p} + u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \cdots + \psi_q u_{t-q}, \tag{1.27}$$

where $\phi_j$and ψ_k are unknown parameters and $u_t \sim N(0, \sigma^2)$.

– Thus under ARMA processes the matrices **V** and **w** are depending on the vector of parameters $\boldsymbol{\theta}$, which contains all the parameters $\phi_j$and ψ_k.

Example 1.1.

Consider the dataset pollutiondata.txt:

```
> data<-read.table("pollutiondata.txt",header=TRUE,sep="\t")
> head(data)
Mortality Temperature Particulates
1 97.85 72.38 72.72
2 104.64 67.19 49.60
3 94.36 62.94 55.68
4 98.05 72.49 55.16
5 95.85 74.25 66.02
6 95.98 67.88 44.01
.
.
.
507  89.43 73.33 57.58
508  NA  70.52 62.61
```

Model the variable $Y = \mathsf{Mortality}$ by the linear model

$$y_t = \beta_0 + \beta_1 x_{(t-1)1} + \beta_2 x_{(t-1)2} + \varepsilon_t,$$

where $X_1 = \mathsf{Temperature}$ and $X_2 = \mathsf{Particulates}$. Consider different ARMA processes

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \cdots + \phi_p \varepsilon_{t-p} + u_i + \psi_1 u_{i-1} + \psi_2 u_{i-2} + \cdots + \psi_q u_{i-q}$$

for the random error term ε_t. After finding suitable model, consider the following questions.

(a) Calculate the (restricted) maximum likelihood estimates for the parameters of the ARMA process.

(b) Test the statistical significance of the variable $X_2 = \mathsf{Particulates}$ in the model.

(c) Consider predicting the value of the new observation y_{t_*} at the last time point $t_* = 508$. Calculate the (empirical) best linear unbiased prediction for the y_{t_*}.

(d) Consider predicting the value of the new observation y_{t_*} at the last time point $t_* = 508$. Construct 80% prediction interval for the y_{t_*}.

1.2 Modeling with Linear Mixed Models

1.2.1 Types of Linear Mixed Effects Models

– In linear mixed models, there are two set of explanatory variables $X_1, X_2, \ldots, X_p$ and $Z_1, Z_2, \ldots, Z_q$ which are effecting linearly the expected value μ of the normally distributed response variable Y :

$$\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + b_1 Z_1 + b_2 Z_2 + \cdots + b_q Z_q, \tag{1.28}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are unknown fixed parameters, and $b_1, \ldots, b_q$ are unknown *random effects*.

– Each random effect $b_1, b_2, \ldots, b_q$ is assumed to follow normal distribution

$$b_k \sim N(0, \sigma^2_{z_k}), \quad k = 1, 2, \ldots, q. \tag{1.29}$$

– If the explanatory variables $Z_1, Z_2, \ldots, Z_q$ are dummy variables defining the categories of some underlying nominal variable, then the linear mixed model is often called as linear *variance component model*, and it has the form

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + b_{z_{i1}} + b_{z_{i2}} + \cdots + b_{z_{iq}} + \varepsilon_i, \tag{1.30}$$

where $b_{z_{i1}}, b_{z_{i2}}, \ldots, b_{z_{iq}}$ are random effects associated to the dummy variables $Z_1, Z_2, \ldots, Z_q$, respectively, and ε_i is random error.

– The random effects $b_{z_{ik}}$ are assumed to follow the normal distribution

$$b_{z_{ik}} \sim N(0, \sigma^2_{z_k}),$$

where $\sigma^2_{z_k}$ is unknown parameter which needs to be estimated by the data.

– The random noise ε_i is assumed to follow the normal distribution

$$\varepsilon_i \sim N(0, \sigma^2),$$

where σ^2 is unknown parameter which needs to be estimated by the data.

– The name of the linear variance component model follows from the fact that

$$\begin{aligned}\mathrm{Var}(Y_i) &= \mathrm{Var}(b_{z_{i1}}) + \mathrm{Var}(b_{z_{i2}}) + \cdots + \mathrm{Var}(b_{z_{iq}}) + \mathrm{Var}(\varepsilon_i) \\ &= \sigma^2_{z_1} + \sigma^2_{z_2} + \cdots + \sigma^2_{z_q} + \sigma^2.\end{aligned}$$

– Often from the same sampling unit i, it is measured repeatedly with respect to time or space, T, the value of the response variable Y.

– In repeated measures situation, the considered random variable is Y_{it}, where index i defines the sampling unit and index t the time or space when the random variable occurs to the sampling unit i.

– In linear mixed models for repeated measures, the random response variable Y_{it} is considered to have a form

$$Y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_p x_{itp} + b_{i0} + b_{i1} z_{it1} + b_{i2} z_{it2} + \cdots + b_{iq} z_{itq} + \varepsilon_{it},$$

where $b_{i0}, b_{i1}, b_{i2}, \ldots, b_{iq}$ are random effects following normal distributions

$$b_{i0} \sim N(0, \sigma^2_{b_0}),\ b_{i1} \sim N(0, \sigma^2_{b_1}), \ldots, b_{iq} \sim N(0, \sigma^2_{b_q}).$$

– Linear mixed models often contains random effects related to repeated measures and random effects related to some explanatory variables $Z_1, Z_2, \ldots, Z_q$.

– Usually, a linear mixed model can be written in form

$$\begin{aligned} \mathcal{M}_{X_{1|2|\ldots|p}|Z_{1|2|\ldots|q}}: \quad Y_{it} =& \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \cdots + \beta_p x_{itp} \\ &+ b_{i0} + b_{i1} z_{it1} + b_{i2} z_{it2} + \cdots + b_{iq} z_{itq} \\ &+ b_{z_{it(q+1)}} + b_{z_{it(q+2)}} + \cdots + b_{z_{itr}} + \varepsilon_{it}. \end{aligned}$$

– Linear mixed effects model for the sampling units $i = 1, 2, \ldots, n$ can be expressed in matrix form as

$$\mathbf{y}|\mathbf{b} \sim N(\boldsymbol{\mu}_{\mathbf{b}}, \sigma^2 \mathbf{I}), \tag{1.31a}$$

$$\boldsymbol{\mu}_{\mathbf{b}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \tag{1.31b}$$

$$\mathbf{b} \sim N(\mathbf{0}, \mathbf{G}), \tag{1.31c}$$

where the covariance matrix $\mathbf{G}$ is often the block matrix associated to the partitioned $\mathbf{Z} = (\mathbf{Z}_1 : \mathbf{Z}_2 : \cdots : \mathbf{Z}_q)$ and $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_q)$:

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 & \ldots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{G}_q \end{pmatrix}.$$

– Under the linear mixed effects model, the marginal distribution of $\mathbf{y}$ is

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \qquad \mathbf{V} = \sigma^2\mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}', \tag{1.32}$$

and the joint distribution of $\mathbf{y}$ and $\mathbf{b}$ is

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{b} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{Z}\mathbf{G} \\ \mathbf{G}'\mathbf{Z}' & \mathbf{G} \end{pmatrix}\right]. \tag{1.33}$$

1.2.2 Longitudinal Mixed Effects Models

– In repeated measures or in longitudinal data situation, for each sampling unit i, there is an observable random vector

$$\mathbf{y}_i = \{y_{it}\} = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \qquad t = 1, 2, \ldots, T.$$

– For each sampling unit i, the linear mixed effects model has the form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, 2, \ldots, N, \tag{1.34}$$

with assumptions of

$$\mathrm{Cov}(\mathbf{b}_i) = \mathbf{F}, \qquad \mathrm{Cov}(\boldsymbol{\varepsilon}_i) = \sigma^2\mathbf{I}, \qquad \mathrm{Cov}(\mathbf{b}_i, \boldsymbol{\varepsilon}_i) = \mathbf{0}. \tag{1.35}$$

– Thus the marginal model is

$$\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i), \qquad \mathbf{V}_i = \sigma^2\mathbf{I} + \mathbf{Z}_i\mathbf{F}\mathbf{Z}_i'. \tag{1.36}$$

– The model for the whole data can be written as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \qquad \mathrm{Cov}(\mathbf{b}) = \mathbf{G} = \mathbf{I} \otimes \mathbf{F}, \\ \mathbf{y} &\sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \qquad \mathbf{V} = \sigma^2\mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}', \end{aligned} \tag{1.37}$$

where

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_N \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix}.$$

– In lme4 library, the function lmer writes the individual model in a form

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \qquad \mathrm{Cov}(\mathbf{b}_i) = \sigma^2\mathbf{F}, \tag{1.38}$$

and the general model in a form

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \qquad \mathrm{Cov}(\mathbf{b}) = \sigma^2\mathbf{G} = \sigma^2\mathbf{I} \otimes \mathbf{F}, \\ \mathbf{y} &\sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}), \qquad \sigma^2\mathbf{V} = \sigma^2\mathbf{I} + \sigma^2\mathbf{Z}\mathbf{G}\mathbf{Z}. \end{aligned} \tag{1.39}$$

Example 1.2.

The aim of the study was to find out how the salinity of the soil affects the amount of ash content of the saltgrass Distichlis spicata when salinity experiments were carried out in 16 different test areas. The material below is found in saltgrass.txt.

```
    Ashcontent Salinity Location
1        61.31        1        1
2        56.68        1        1
3        63.89        1        1
.
384      89.52        4       16

Description: Ash content (g/kg DM) of saltgrass
for 16 locations at 4 salinity
levels (1.5, 10, 30, 50 dS/m),
with 6 replicates per treatment combination.

Z=Location    Y=Ashcontent
X=Salinity     1=1.5, 2=10, 3=30, 4=50
```

Denote the variables as the following: $Y =$ Ashcontent, $X =$ Salinity, with index j associated to it, $Z =$ Location, with index h associated to it. Consider the linear variance component model

$$\mathcal{M}: \quad y_i = \beta_0 + \beta_j + b_h + \varepsilon_i.$$

where β_0, β_j are fixed parameters associated to the variable $X =$ Salinity, and b_h are random effects parameters associated to the variable $Z =$ Location with the assumption of $b_h \sim N(0, \sigma_z^2)$. Furthermore $\varepsilon_i \sim N(0, \sigma^2)$. Write in details the model $\mathcal{M}$ in matrix form

$$\begin{aligned} \mathbf{y}|\mathbf{b} &\sim N(\boldsymbol{\mu}_\mathrm{b}, \sigma^2\mathbf{I}), \\ \boldsymbol{\mu}_\mathrm{b} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}, \\ \mathbf{b} &\sim N(\mathbf{0}, \mathbf{G}). \end{aligned}$$

Example 1.3.

Consider the dataset fertilizer.txt:

```
    root week plant fertilizer
1   1.30    2   ID1      added
2   3.50    4   ID1      added
3   7.00    6   ID1      added
4   8.10    8   ID1      added
5  10.00   10   ID1      added
6   2.00    2   ID2      added
7   3.50    4   ID2      added
.
.

A completely randomized design with a single factor
(with fertilizer or without). Each treatment is applied
to six plants and the root length of each plant
is measured on five occasions (week 2, 4, 6, 8, 10).
The response variable is the root length.
```

Denote variables as following: Y=root, X_1=fertlizer, X_2=week(=T). Consider the following linear mixed effects model

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it1} x_{it2} + b_{i0} + b_{i1} x_{it2} + \varepsilon_{it},$$

where x_{it1} is dummy variable having the values 0 and 1, and actually $x_{it2} = t_i$.

Write in details the model $\mathcal{M}$ in matrix form

$$\mathbf{y}|\mathbf{b} \sim N(\boldsymbol{\mu}_{\mathrm{b}}, \sigma^2\mathbf{I}),$$
$$\boldsymbol{\mu}_{\mathrm{b}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b},$$
$$\mathbf{b} \sim N(\mathbf{0}, \mathbf{G}).$$

1.2.3 Estimation in Linear Mixed Effects Models

– In linear mixed effects models, the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ and the maximum likelihood predictor $\hat{\mathbf{b}}$ are obtained by jointly maximizing the log-likelihood function of the joint distribution

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{b} \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{ZG} \\ \mathbf{GZ}' & \mathbf{G} \end{pmatrix}\right], \tag{1.40}$$

i.e., $\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}} = \arg\max_{\boldsymbol{\beta},\mathbf{b}} l_{\mathbf{y},\mathbf{b}}(\boldsymbol{\beta}, \mathbf{b}|\mathbf{y})$.

– Setting $\frac{\partial l_{\mathbf{y},\mathbf{b}}(\boldsymbol{\beta},\mathbf{b}|\mathbf{y})}{\partial \boldsymbol{\beta}} = \mathbf{0}$, $\frac{\partial l_{\mathbf{y},\mathbf{b}}(\boldsymbol{\beta},\mathbf{b}|\mathbf{y})}{\partial \mathbf{b}} = \mathbf{0}$, then $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$ must be solutions to the Henderson's mixed model equations

$$\begin{pmatrix} \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} & \frac{1}{\sigma^2}\mathbf{X}'\mathbf{Z} \\ \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{X} & \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}\mathbf{X}'\mathbf{y} \\ \frac{1}{\sigma^2}\mathbf{Z}'\mathbf{y} \end{pmatrix}. \tag{1.41}$$

– Solutions for the Henderson's mixed model equations are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{1.42a}$$

$$\hat{\mathbf{b}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{1.42b}$$

where in final expressions the covariance matrices $\mathbf{V}$ and $\mathbf{G}$ are replaced with estimates $\widetilde{\mathbf{V}}$ and $\widetilde{\mathbf{G}}$.

– For the given value of $\mathbf{x}_{i*}$, the estimator

$$\hat{\mu}_{i*} = \mathbf{x}'_{i*}\hat{\boldsymbol{\beta}} = \mathbf{x}'_{i*}(\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{y} \tag{1.43}$$

is called the maximum likelihood estimator of the μ_{i*}.

– For the given values of $\mathbf{x}_{i*}$ and $\mathbf{z}_{i*}$, the predictor

$$\tilde{\mu}_{i*} = \mathbf{x}'_{i*}\hat{\boldsymbol{\beta}} + \mathbf{z}'_{i*}\hat{\mathbf{b}} = \mathbf{x}'_{i*}(\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{y} + \mathbf{z}'_{i*}\widetilde{\mathbf{G}}\mathbf{Z}'\widetilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \tag{1.44}$$

is called the maximum likelihood predictor of the μ_{i*}.

– Let us denote by the vector $\boldsymbol{\sigma}^2_{\mathrm{b}}$ unknown parameters of the matrix $\mathbf{G} = \mathbf{G}_{\boldsymbol{\sigma}^2_{\mathrm{b}}}$.

– The parameters σ^2 and $\boldsymbol{\sigma}^2_{\mathrm{b}}$ are usually estimated by the restricted maximum likelihood methods, which means that estimators $\tilde{\sigma}^2$ and $\tilde{\boldsymbol{\sigma}}^2_{\mathrm{b}}$ are maximizing the log-likelihood related to the residual distribution

$$\mathbf{L}'\mathbf{y} \sim N\left(\mathbf{0}, \mathbf{L}'(\sigma^2\mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}')\mathbf{L}\right), \tag{1.45}$$

where $\mathbf{L}$ such matrix that $\mathbf{L}'\mathbf{X} = \mathbf{0}$ and for the columnspaces $\mathcal{C}(\mathbf{L}) = \mathcal{C}(\mathbf{X})^{\perp}$, i.e.,

$$\tilde{\sigma}^2, \tilde{\boldsymbol{\sigma}}^2_{\mathbf{b}} = \arg\max_{\sigma^2, \boldsymbol{\sigma}^2_{\mathbf{b}}} l_{\mathbf{L}'\mathbf{y}}(\sigma^2, \boldsymbol{\sigma}^2_{\mathbf{b}} | \mathbf{L}'\mathbf{y}). \tag{1.46}$$

1.2.4 Testing in Linear Mixed Effects Models

– Consider two hierarchical mixed effects models

$$\begin{aligned} \mathcal{M}_1 : \quad & \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \\ \mathcal{M}_{1|2} : \quad & \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \end{aligned}$$

where $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$.

– The models have marginal distributions

$$\begin{aligned} \mathcal{M}_1 : \quad & \mathbf{y} \sim N\left(\mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{V}\right), \\ \mathcal{M}_{1|2} : \quad & \mathbf{y} \sim N\left(\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\right), \end{aligned}$$

where $\mathbf{V} = \sigma^2\mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$.

– Consider the hypotheses

$$\begin{aligned} H_0 : & \text{Model } \mathcal{M}_1 \text{ is the true model}, \\ H_1 : & \text{Model } \mathcal{M}_{1|2} \text{ is the true model}. \end{aligned}$$

– Likelihood ratio statistic based on marginal distributions

$$LR = -2 \cdot \left(l_{\mathcal{M}_1}(\hat{\boldsymbol{\beta}}_1, \hat{\sigma}^2, \hat{\boldsymbol{\sigma}}^2_{\mathrm{b}}|\mathbf{y}) - l_{\mathcal{M}_{1|2}}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \hat{\sigma}^2, \hat{\boldsymbol{\sigma}}^2_{\mathrm{b}}|\mathbf{y}) \right)$$

follows asymptotically χ^2 distribution with degrees of freedom $\mathrm{rank}(\mathbf{X}_2)$.

– Consider the linear hypotheses

$$H_0 : \mathbf{K}'\boldsymbol{\beta} = \mathbf{0},$$
$$H_1 : \mathbf{K}'\boldsymbol{\beta} \neq \mathbf{0}, \qquad \mathbf{K}' \in \mathbb{R}^{q,(p+1)}.$$

– Wald statistic

$$W = (\mathbf{K}'\hat{\boldsymbol{\beta}})'(\mathbf{K}'(\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{K})^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}} = (\mathbf{K}'\hat{\boldsymbol{\beta}})'\left(\widetilde{\mathrm{Cov}}(\mathbf{K}'\hat{\boldsymbol{\beta}})\right)^{-1}\mathbf{K}'\hat{\boldsymbol{\beta}} \tag{1.47}$$

follows asymptotically χ^2 distribution with degrees of freedom $\mathrm{rank}(\mathbf{K})$.

– Consider two residual distributions

$$\mathcal{M}_1 : \quad \mathbf{L}'\mathbf{y} \sim N\left(\mathbf{0}, \mathbf{L}'(\sigma^2\mathbf{I} + \mathbf{Z}\mathbf{G}(\boldsymbol{\sigma}^2_{1_\mathbf{b}})\mathbf{Z}')\mathbf{L}\right),$$
$$\mathcal{M}_{1|2} : \quad \mathbf{L}'\mathbf{y} \sim N\left(\mathbf{0}, \mathbf{L}'(\sigma^2\mathbf{I} + \mathbf{Z}\mathbf{G}(\boldsymbol{\sigma}^2_{1_\mathbf{b}}, \boldsymbol{\sigma}^2_{2_\mathbf{b}})\mathbf{Z}')\mathbf{L}\right),$$

related to the hypotheses $\boldsymbol{\sigma}^2_{\mathbf{b}} = (\boldsymbol{\sigma}^2_{1_\mathbf{b}}, \boldsymbol{\sigma}^2_{2_\mathbf{b}})'$

$$H_0 : \boldsymbol{\sigma}^2_{2_\mathbf{b}} = \mathbf{0},$$
$$H_1 : \boldsymbol{\sigma}^2_{2_\mathbf{b}} \neq \mathbf{0}.$$

– Likelihood ratio statistic based on residual distributions

$$LR = -2 \cdot \left(l_{\mathcal{M}_1}(\tilde{\sigma}^2, \tilde{\boldsymbol{\sigma}}^2_{1_\mathbf{b}}|\mathbf{L}'\mathbf{y}) - l_{\mathcal{M}_{1|2}}(\tilde{\sigma}^2, \tilde{\boldsymbol{\sigma}}^2_{1_\mathbf{b}}, \tilde{\boldsymbol{\sigma}}^2_{2_\mathbf{b}}|\mathbf{L}'\mathbf{y})\right)$$

follows (almost) asymptotically χ^2 distribution with degrees of freedom being number of parameters in the vector $\boldsymbol{\sigma}^2_{2_\mathbf{b}}$.

1.2.5 Prediction in Linear Mixed Effects Models

– In case of linear mixed models, consider the prediction of the new observation y_f (observable in future) with given values of the explanatory variables $\mathbf{x}_f$ and $\mathbf{z}_f$:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \tag{1.48a}$$

$$y_f = \mathbf{x}_f'\boldsymbol{\beta} + \mathbf{z}_f'\mathbf{b} + \varepsilon_f \tag{1.48b}$$

– If $\mathbf{z}_f$ is such that $\mathbf{z}_f \in \mathcal{C}(\mathbf{Z}')$, the maximum likelihood predictor (also the BLUP) for the new observation is

$$\hat{y}_f = \mathbf{x}_f'\hat{\boldsymbol{\beta}} + \mathbf{z}_f'\hat{\mathbf{b}} = \mathbf{x}_f'(\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{y} + \mathbf{z}_f'\widetilde{\mathbf{G}}\mathbf{Z}'\widetilde{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \tag{1.49}$$

– Otherwise, the maximum likelihood predictor (also the BLUP) for the new observation is $\hat{y}_f = \mathbf{x}_f'\hat{\boldsymbol{\beta}} = \mathbf{x}_f'(\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widetilde{\mathbf{V}}^{-1}\mathbf{y}$.

– Note that joint marginal distribution of $\mathbf{y}$ and y_f is

$$\begin{pmatrix} \mathbf{y} \\ y_f \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{x}_f'\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{Z}\mathbf{G}\mathbf{z}_f \\ \mathbf{z}_f'\mathbf{G}\mathbf{Z}' & \sigma^2 + \mathbf{z}_f'\mathbf{G}\mathbf{z}_f \end{pmatrix}\right], \tag{1.50}$$

and hence for the BLUP of y_f, the variance of the prediction error is

$$\begin{aligned} \operatorname{Var}(y_f - \hat{y}_f) &= \sigma^2\Sigma_{e_f} \\ &= v_f - \mathbf{w}'\mathbf{V}^{-1}\mathbf{w} + (\mathbf{x}_f' - \mathbf{w}'\mathbf{V}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{x}_f - \mathbf{X}'\mathbf{V}^{-1}\mathbf{w}), \end{aligned} \tag{1.51}$$

where $v_f = \sigma^2 + \mathbf{z}_f'\mathbf{G}\mathbf{z}_f$ and $\mathbf{w} = \mathbf{z}_f'\mathbf{G}\mathbf{Z}'$.

A. Asymptotic prediction error based method

A1. Consider the estimate of variance of the prediction error $e_f = y_f - \hat{y}_f$, i.e, find $\widetilde{\mathrm{Var}}(y_f - \hat{y}_f) = \tilde{\sigma}^2_{e_f}$.

A2. Assume that asymptotically $y_f - \hat{y}_f \sim N\left(0, \tilde{\sigma}^2_{e_f}\right)$.

A3. Construct the $100(1-\alpha)\%$ prediction interval

$$\left[\hat{y}_f - z_{\alpha/2}\sqrt{\tilde{\sigma}^2_{e_f}}, \hat{y}_f + z_{\alpha/2}\sqrt{\tilde{\sigma}^2_{e_f}}\right].$$

B. Parametric bootstrap based method

B1. Find the predictor $\tilde{\mu}_f = \mathbf{x}'_f\hat{\boldsymbol{\beta}} + \mathbf{z}'_f\hat{\mathbf{b}}_f$

B2. Simulate $\tilde{\mu}_{f_*}$ from the distribution $\tilde{\mu}_{f_*} \sim N(\tilde{\mu}_f, \widetilde{\mathrm{Var}}(\tilde{\mu}_f))$.

B3. Simulate y_{f_*} from the distribution $y_{f_*} \sim N(\tilde{\mu}_{f_*}, \widetilde{\mathrm{Var}}(\varepsilon_f))$.

B4. Repeat M times the steps B2-B3, and then determine $\alpha/2$ and $1-\alpha/2$ the quantiles of the simulated values y_{f_*} .

C. Estimated quantiles method

C1. Find the predictor $\tilde{\mu}_f = \mathbf{x}'_f\hat{\boldsymbol{\beta}} + \mathbf{z}'_f\hat{\mathbf{b}}_f$, and the estimate $\widetilde{\mathrm{Var}}(\varepsilon_f)$.

C2. Determine $\alpha/2$ and $1-\alpha/2$ the quantiles of the estimated distribution $y_f \sim N(\tilde{\mu}_f, \widetilde{\mathrm{Var}}(\varepsilon_f))$.

Example 1.4.

The aim of the study was to find out how the salinity of the soil affects the amount of ash content of the saltgrass Distichlis spicata when salinity experiments were carried out in 16 different test areas. The material below is found in saltgrass.txt.

```
    Ashcontent Salinity Location
1        61.31        1        1
2        56.68        1        1
3        63.89        1        1
4        68.78        1        1
5        68.41        1        1
6        70.94        1        1
7        65.99        2        1
.
382      87.38        4       16
383      84.79        4       16
384      89.52        4       16

Description: Ash content (g/kg DM) of saltgrass
for 16 locations at 4 salinity
levels (1.5, 10, 30, 50 dS/m),
with 6 replicates per treatment combination.

Variables:
Z=Location
X=Salinity     1=1.5, 2=10, 3=30, 4=50  */
Y=Ashcontent
```

Denote the variables as the following: $Y =$ Ashcontent, $X =$ Salinity, with index j associated to it, $Z =$ Location, with index h associated to it. Consider the linear variance component model

$$\mathcal{M}: \quad y_i = \beta_0 + \beta_j + b_h + \varepsilon_i.$$

where β_0, β_j are fixed parameters associated to the variable $X =$ Salinity, and b_h are random effects parameters associated to the variable $Z =$ Location with the assumption of $b_h \sim N(0, \sigma_z^2)$. Furthermore $\varepsilon_i \sim N(0, \sigma^2)$.

(a) Under the model $\mathcal{M}$, calculate (restricted) maximum likelihood estimate for the parameter σ_z^2.

(b) For which salinity level, the expected value μ_{i*} is estimated to highest?

   (a) 1=1.5 dS/m,

   (b) 2=10 dS/m,

   (c) 3=30 dS/m,

   (d) 4=50 dS/m.

(c) Under the model $\mathcal{M}$, calculate the maximum likelihood prediction $\hat{b}_{12}$ for the random effect b_{12} related to value $z_{i*} = 12$.

(d) Test at 5% significance level, is the explanatory variable $X = \mathsf{Salinity}$ statistically significant variable under the model $\mathcal{M}$. Calculate the value of the test statistic.

(e) Under the model $\mathcal{M}$, calculate the maximum likelihood prediction $\tilde{\mu}_{i*}$ for the expected value μ_{i*} when $x_{i*} = 2 = 10$ dS/m and $z_{i*} = 5$.

(f) Under the model $\mathcal{M}$, construct 80% prediction interval for the new observation y_f when $x_f = 2 = 10$ dS/m and $z_f = 5$.

Example 1.5.

Consider the dataset fertilizer.txt:

```
    root week plant fertilizer
1    1.30    2   ID1      added
2    3.50    4   ID1      added
3    7.00    6   ID1      added
4    8.10    8   ID1      added
5   10.00   10   ID1      added
6    2.00    2   ID2      added
7    3.50    4   ID2      added
8    5.50    6   ID2      added
9    7.20    8   ID2      added
10   9.10   10   ID2      added
11   1.70    2   ID3      added
12   3.20    4   ID3      added
13   5.80    6   ID3      added
.
.
```

```
A completely randomized design with a
single factor (with fertilizer or without).
Each treatment is applied to six plants
and the root length of each plant
is measured on five occasions
(week 2, 4, 6, 8, 10).
The response variable is the root length.
```

Denote variables as following: Y=root, X_1=fertlizer, X_2=week(=T).

Consider the following linear mixed effects model

$$\mathcal{M}: \quad y_{it} = \beta_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + \beta_3 x_{it1} x_{it2} + b_{i0} + b_{i1} x_{it2} + \varepsilon_{it},$$

where x_{it1} is dummy variable having the values 0 and 1, and actually $x_{it2} = t_i$.

(a) Under the model $\mathcal{M}$, calculate the maximum likelihood predictions for the random effects b_{i1}.

(b) Under the model $\mathcal{M}$, calculate the (restricted) maximum likelihood estimate for the covariance matrix

$$\mathrm{Cov}\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} = \mathbf{F}.$$

(c) Test at 5% significance level, is the explanatory variable $X_1 = \mathsf{fertlizer}$ statistically significant variable in the model $\mathcal{M}$. Calculate the value of the test statistic.

(d) Construct 80 % prediction interval for new observation y_f for the plant ID1 to the time point $\mathsf{week} = 11$.

Example 1.6.

Consider the following dataset:

```
> interaction.plot(week, plant,root)
> library(faraway)

> data(psid)
> psid
    age educ sex income year person
1    31   12   M   6000   68      1
2    31   12   M   5300   69      1
3    31   12   M   5200   70      1
4    31   12   M   6900   71      1
5    31   12   M   7500   72      1
6    31   12   M   8000   73      1
.
```

Denote variables as following: Y=income, T=year, X=sex.

Model the data by the following linear mixed effects model

$$\mathcal{M}: \quad y_{it} = \beta_0 + \beta_1 t_i + \alpha_j + b_{i0} + b_{i1} t_i + \varepsilon_{it},$$

where α_j are fixed parameters related to the categories of the variable X=sex.

(a) Under the model $\mathcal{M}$, calculate the (restricted) maximum likelihood estimate for the covariance matrix

$$\mathrm{Cov}\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} = \mathbf{F}.$$

(b) Construct 80 % prediction interval for new observation y_f for the person 1 to the time point $T = 84$.