Lecture 6 | 15.04.2024

Parametric structures for variance/covariance in mixed models

Some overview

- Consider a linear regression model for repeated measurements within (independent) subjects $i \in \{1, \dots, N\}$ in a form

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

where $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ is the subject specific response vector, $\mathbb{X}_i = (\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i})^\top$ for $\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijp})^\top$ and $j = 1, \dots, n_i$ are the subject (and time specific) explanatory vectors (of dimension $p \in \mathbb{N}$) and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the unknown vector of the regression parameters (mean structure) same for all subjects and time points $\boldsymbol{t}_i = (t_{i1}, \dots, t_{in_i})^\top$

- the variance-covariance structure within each subject is modelled by the vector parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$, such that $\varepsilon_i \sim N_{n_i}(\mathbf{0}_i, \mathbb{V}_i(\boldsymbol{t}_i, \boldsymbol{\alpha}))$
- the stochastic (non-systematic) term of the model – random errors ε_{ij} are decomposed into three main parts: the random effects, the serial correlation, and the measurement error

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

Some overview

- Consider a linear regression model for repeated measurements within (independent) subjects $i \in \{1, \dots, N\}$ in a form

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

where $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ is the subject specific response vector, $\mathbb{X}_i = (\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{in_i})^\top$ for $\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijp})^\top$ and $j = 1, \dots, n_i$ are the subject (and time specific) explanatory vectors (of dimension $p \in \mathbb{N}$) and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is the unknown vector of the regression parameters (mean structure) same for all subjects and time points $\boldsymbol{t}_i = (t_{i1}, \dots, t_{in_i})^\top$

- the variance-covariance structure within each subject is modelled by the vector parameters $\boldsymbol{\alpha} \in \mathbb{R}^q$, such that $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\boldsymbol{0}_i, \mathbb{V}_i(\boldsymbol{t}_i, \boldsymbol{\alpha}))$
- the stochastic (non-systematic) term of the model – random errors ε_{ij} are decomposed into three main parts: the random effects, the serial correlation, and the measurement error

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

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

Stochastic properties of the error terms

- measurement errors $\omega_{ij} \sim N(0, \tau^2)$, mutually independent for i and j
    - lets denote $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{in_i})^\top$ and $\boldsymbol{\omega}_i \sim N_{n_i}(\mathbf{0}, \tau^2 \mathbb{I}_i)$, for $\mathbb{I}_i \in \mathbb{R}^{n_i \times n_i}$

Stochastic properties of the error terms

- measurement errors $\omega_{ij} \sim N(0, \tau^2)$, mutually independent for i and j
  - lets denote $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{in_i})^\top$ and $\boldsymbol{\omega}_i \sim N_{n_i}(\mathbf{0}, \tau^2 \mathbb{I}_i)$, for $\mathbb{I}_i \in \mathbb{R}^{n_i \times n_i}$

- serial correlation represented by random variables $W_i(t_{ij})$ sampled from $N \in \mathbb{N}$ independent copies of a stationary Gaussian process $\{W(t);\ t \in \mathbb{R}\}$, with the zero mean, variance $\sigma^2 > 0$ and the correlation function $\rho(u) = cor(W(t), W(t+u))$
  - random variables $W_{ij} \equiv W_i(t_{ij})$ are independent with respect to subjects $i \in \{1, \dots, N\}$ but dependent within subjects, i.e., for indexes $j = 1, \dots, n_i$
  - lets denote $\mathbb{H}_i = (h_{ijk})_{j,k=1}^{n_i}$, where $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$, i.e., the correlation between Y_{ij} and Y_{ik}, taken at the time points t_{ij} and t_{ik}
  - thus, for the vector $\boldsymbol{W}_i = (W_{i1}, \dots, W_{in_i})^\top$ we have $Var\,\boldsymbol{W}_i = \sigma^2 \mathbb{H}_i$

Stochastic properties of the error terms

- measurement errors $\omega_{ij} \sim N(0, \tau^2)$, mutually independent for i and j
    - lets denote $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{in_i})^\top$ and $\boldsymbol{\omega}_i \sim N_{n_i}(\mathbf{0}, \tau^2 \mathbb{I}_i)$, for $\mathbb{I}_i \in \mathbb{R}^{n_i \times n_i}$

- serial correlation represented by random variables $W_i(t_{ij})$ sampled from $N \in \mathbb{N}$ independent copies of a stationary Gaussian process $\{W(t);\ t \in \mathbb{R}\}$, with the zero mean, variance $\sigma^2 > 0$ and the correlation function $\rho(u) = cor(W(t), W(t+u))$
    - random variables $W_{ij} \equiv W_i(t_{ij})$ are independent with respect to subjects $i \in \{1, \dots, N\}$ but dependent within subjects, i.e., for indexes $j = 1, \dots, n_i$
    - lets denote $\mathbb{H}_i = (h_{ijk})_{j,k=1}^{n_i}$, where $h_{ijk} = \rho(|t_{ij} - t_{ik}|)$, i.e., the correlation between Y_{ij} and Y_{ik}, taken at the time points t_{ij} and t_{ik}
    - thus, for the vector $\boldsymbol{W}_i = (W_{i1}, \dots, W_{in_i})^\top$ we have $Var\,\boldsymbol{W}_i = \sigma^2 \mathbb{H}_i$

- normally distributed random effects $\boldsymbol{w}_i \sim N_r(\mathbf{0}, \mathbb{G})$, mutually independent for $i = 1, \dots, N$, with the corresponding explanatory variables $\boldsymbol{z}_{ij} \in \mathbb{R}^r$
    - the random effect $\boldsymbol{w}_i$ is only subject specific (index i) but the explanatory vectors $\boldsymbol{z}_{ij}$ related to this random effects are subject and time specific
    - lets denote $\mathbb{Z}_i = (\boldsymbol{z}_{i1}, \dots, \boldsymbol{z}_{in_i})^\top \in \mathbb{R}^{n_i \times r}$

Parametric models for variance/covariance

Variance/covariance of $\boldsymbol{Y}_i \in \mathbb{R}^{n_i}$ can be expressed as

- $Var\,\boldsymbol{Y}_i = Var(\varepsilon_i) = Var\left[\boldsymbol{z}_{ij}^\top \boldsymbol{w}_i + W_i(t_{ij}) + \omega_{ij}\right] = \mathbb{Z}_i \mathbb{G} \mathbb{Z}_i^\top + \underbrace{\sigma^2 \mathbb{H}_i + \tau^2 \mathbb{I}_i}_{\mathbb{R}_i \text{ in SAS}}$

  $\hookrightarrow$ because random quantities $\boldsymbol{w}_i$, $\boldsymbol{W}_i$, and $\boldsymbol{\omega}_i$ are mutually independent

Parametric models for variance/covariance

Variance/covariance of $\boldsymbol{Y}_i \in \mathbb{R}^{n_i}$ can be expressed as

- $Var\,\boldsymbol{Y}_i = Var(\varepsilon_i) = Var\left[\boldsymbol{z}_{ij}^\top \boldsymbol{w}_i + W_i(t_{ij}) + \omega_{ij}\right] = \mathbb{Z}_i \mathbb{G} \mathbb{Z}_i^\top + \underbrace{\sigma^2 \mathbb{H}_i + \tau^2 \mathbb{I}_i}_{\mathbb{R}_i \text{ in SAS}}$

  $\hookrightarrow$ because random quantities $\boldsymbol{w}_i$, $\boldsymbol{W}_i$, and $\boldsymbol{\omega}_i$ are mutually independent

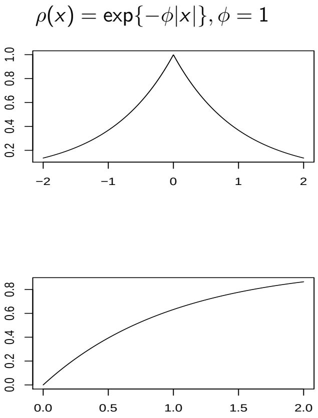
- Thus, the mean structure is fully modelled by the specification of the model matrix $\mathbb{X}_i$ and the vector of parameters $\beta \in \mathbb{R}^p$ but the variance-covariance structure is more complex and it is fully specified by matrices $\mathbb{G}, \mathbb{Z}_i$ and $\mathbb{H}_i$ and, in addition, two parameters $\sigma^2, \tau^2 > 0$
- As the subjects $i \in \{1, \dots, N\}$ are independent, we will only investigate different forms for the variance structure in $Var\,\boldsymbol{Y}_i$, or $Var(\varepsilon_i)$ respectively, for some generic subject $\boldsymbol{Y} \in \mathbb{R}^n$, with $n \in \mathbb{N}$ repeated measurements taken at the time points at $\boldsymbol{t} = (t_1, \dots, t_n)^\top \in \mathbb{R}^n$
- the overall variance-covariance structure for $Var\,\boldsymbol{Y}_i$ will be a block-diagonal matrix with squared matrices of the types $n_i \times n_i$ in the diagonal

Example: correlation & variogram

❏ Consider two different (theoretical) correlation functions ρ and the corresponding variogram functions $\gamma(x) = \sigma^2(1 - \rho(x))$, for $\sigma^2 = 1$

Example: correlation & variogram

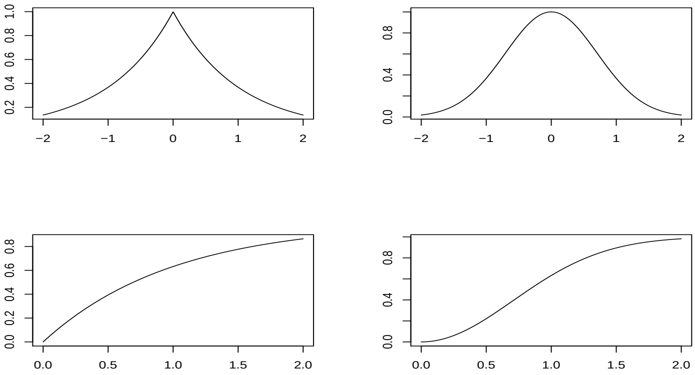
❑ Consider two different (theoretical) correlation functions ρ and the corresponding variogram functions $\gamma(x) = \sigma^2(1 - \rho(x))$, for $\sigma^2 = 1$



Example: correlation & variogram

❑ Consider two different (theoretical) correlation functions ρ and the corresponding variogram functions $\gamma(x) = \sigma^2(1 - \rho(x))$, for $\sigma^2 = 1$

$\rho(x) = \exp\{-\phi|x|\}, \phi = 1$ $\qquad\qquad$ $\rho(x) = \exp\{-\phi x^2\}, \phi = 1$



Exponential model $\longrightarrow$ Gaussian correlation model

Serial correlation (only) model

- from the three possible variance/covariance terms in the expansion

$$Var\,\boldsymbol{Y} = Var(\varepsilon) = \mathbb{Z}\mathbb{G}\mathbb{Z}^{\top} + \sigma^2\mathbb{H} + \tau^2\mathbb{I}$$

  there is only one that is indeed specified—the matrix $\mathbb{H} \in \mathbb{R}^{n\times n}$

Serial correlation (only) model

- from the three possible variance/covariance terms in the expansion

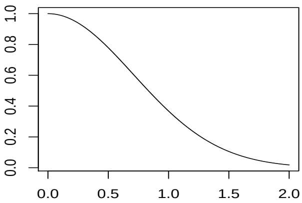
$$Var\,\boldsymbol{Y} = Var(\varepsilon) = \mathbb{Z}\mathbb{G}\mathbb{Z}^\top + \sigma^2\mathbb{H} + \tau^2\mathbb{I}$$

  there is only one that is indeed specified—the matrix $\mathbb{H} \in \mathbb{R}^{n\times n}$
- this implies that $Var(\varepsilon) = \sigma^2\mathbb{H}$ (respectively, $Cov(\varepsilon_j, \varepsilon_k) = \sigma^2\rho(|t_j - t_k|)$ with the corresponding variogram of the form $\gamma(u) = \sigma^2(1 - \rho(u)))$
- all variability is captured within $\sigma^2 > 0$ and the correlation structure by $\mathbb{H}$

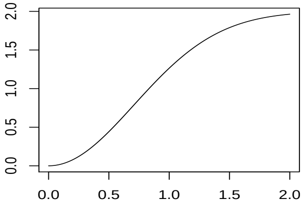
Serial correlation (only) model

- from the three possible variance/covariance terms in the expansion
$$\mathit{Var}\,\boldsymbol{Y} = \mathit{Var}(\varepsilon) = \mathbb{Z}\mathbb{G}\mathbb{Z}^\top + \sigma^2\mathbb{H} + \tau^2\mathbb{I}$$
there is only one that is indeed specified—the matrix $\mathbb{H} \in \mathbb{R}^{n\times n}$
- this implies that $\mathit{Var}(\varepsilon) = \sigma^2\mathbb{H}$ (respectively, $\mathit{Cov}(\varepsilon_j, \varepsilon_k) = \sigma^2\rho(|t_j - t_k|)$ with the corresponding variogram of the form $\gamma(u) = \sigma^2(1 - \rho(u)))$
- all variability is captured within $\sigma^2 > 0$ and the correlation structure by $\mathbb{H}$

$\rho(x) = \exp\{-\phi x^2\}$, $\phi = 1$

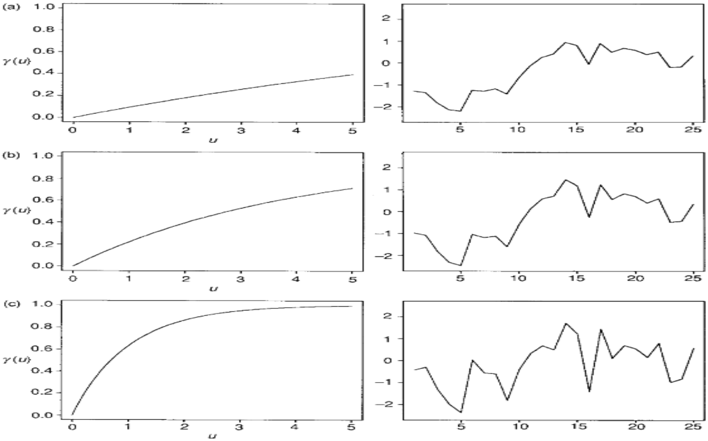


$\gamma(x) = \sigma^2(1 - \rho(x))$, $\sigma^2 = 2$



Parametric models for variance/covariance

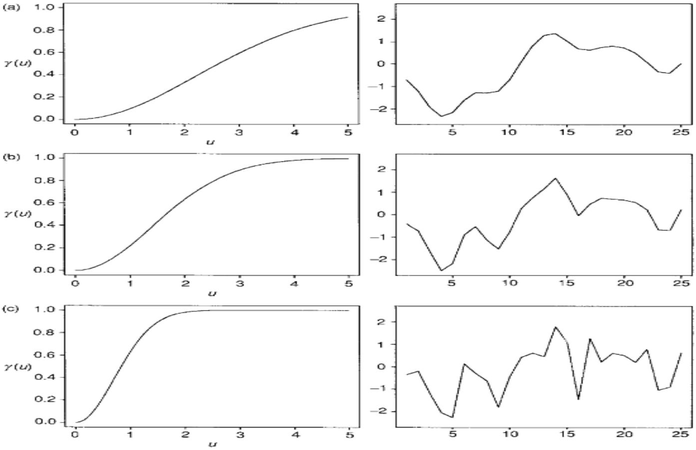
Considering both models—the exponential model and the Gaussian correlation model—the role of the $\phi > 0$ parameter is the same—as the value of ϕ increases, the variograms rises more sharply and the simulated realizations are less smooth (Fig. 5.1., Diggle et al., 2002)



Gaussian correlation model: from top to bottom, $\phi = 0.1$, $\phi = 0.25$, and $\phi = 1.0$

Parametric models for variance/covariance

Considering both models—the exponential model and the Gaussian correlation model—the role of the $\phi > 0$ parameter is the same—as the value of ϕ increases, the variograms rises more sharply and the simulated realizations are less smooth (Fig. 5.2., Diggle et al., 2002)



Exponential model: from top to bottom, $\phi = 0.1$, $\phi = 0.25$, and $\phi = 1.0$

Ante-dependence model

- assuming a direct dependence on previous observations in a sense that ε_j taken at the time point $t_j > t_{j-1} > t_{j-2} > \dots$ depends explicitly on some previous k errors $\varepsilon_{j-1}, \dots, \varepsilon_{j-k}$ *(ante-dependence of order k)*
- also well known as the k-order Markov model or an autoregressive sequence of order k, AR(k)
- AR(1) model formally: $\varepsilon_j = \alpha\varepsilon_{j-1} + \omega_j$, for ω_j being *i.i.d.* from $N(0, \sigma^2)$, for $\varepsilon_0 \sim N(0, \sigma^2/(1-\alpha))$
- the ante-dependence model can be problematic for situations with unequally spaced repeated observations within the subject
- on the other hand, small orders $k \in \mathbb{N}$ can be suitable for the likelihood estimation (straightforward to get the joint distribution/density of ε_i)
- the ante-dependence property is not preserved when incorporating, for instance, the measurements errors

Serial correlation + measurement errors model

- models where there are no random effects present and the variance of ε_i reduces to

$$Var(\varepsilon) = \sigma^2 \mathbb{H} + \tau^2 \mathbb{I}$$

- the variance of ε_j is captured by the sum $\tau^2 + \sigma^2$ with the corresponding variogram function $\gamma(u) = \tau^2 + \sigma^2(1 - \rho(u))$
- the value of $\tau^{>}0$ can be typically estimated from the data that contains duplicate measurements at the same time point

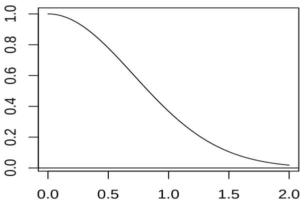
Serial correlation + measurement errors model

- models where there are no random effects present and the variance of ε_i reduces to

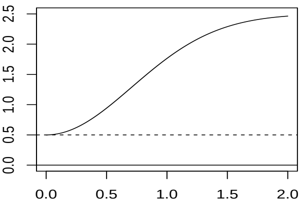
$$Var(\varepsilon) = \sigma^2\mathbb{H} + \tau^2\mathbb{I}$$

- the variance of ε_j is captured by the sum $\tau^2 + \sigma^2$ with the corresponding variogram function $\gamma(u) = \tau^2 + \sigma^2(1 - \rho(u))$
- the value of $\tau^{>}0$ can be typically estimated from the data that contains duplicate measurements at the same time point

$\rho(x) = \exp\{-\phi x^2\}$, $\phi = 1$



$\gamma(x) = \sigma^2(1 - \rho(x))$, $\sigma^2 = 2$, $\tau^2 = 0.5$



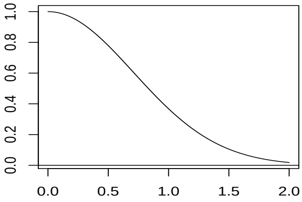
Random intercept model

- the simplest example of a general model with three variance/covariance terms in the decomposition where $z_{ij} = 1$ and $w_i \sim N(0, \nu^2)$
- the variance of ε_j is $Var\varepsilon_j = \nu^2 + \sigma^2 + \tau^2$ and the corelation within the whole vector ε is captured by matrices $\mathbb{H}$ and $\mathbb{J}$ (Diggle, 1988)
- $Var(\varepsilon) = \nu^2\mathbb{J} + \sigma^2\mathbb{H} + \tau^2\mathbb{I}$ with the correlation function $\rho(x)$ and the variogram function $\gamma(u) = \tau^2 + \sigma^2(1 - \rho(u))$
- in this case, however, $\gamma(u)$ does not tend to $Var\varepsilon_j$ as $u \to \infty$

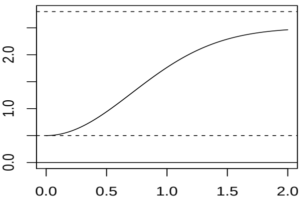
Random intercept model

- the simplest example of a general model with three variance/covariance terms in the decomposition where $z_{ij} = 1$ and $w_i \sim N(0, \nu^2)$
- the variance of ε_j is $Var\varepsilon_j = \nu^2 + \sigma^2 + \tau^2$ and the corelation within the whole vector ε is captured by matrices $\mathbb{H}$ and $\mathbb{J}$ (Diggle, 1988)
- $Var(\varepsilon) = \nu^2\mathbb{J} + \sigma^2\mathbb{H} + \tau^2\mathbb{I}$ with the correlation function $\rho(x)$ and the variogram function $\gamma(u) = \tau^2 + \sigma^2(1 - \rho(u))$
- in this case, however, $\gamma(u)$ does not tend to $Var\varepsilon_j$ as $u \to \infty$

$\rho(x) = \exp\{-\phi x^2\}$, $\phi = 1$



$\gamma(x)$, $\sigma^2 = 2$, $\tau^2 = 0.5$, $\nu^2 = 0.3$



Random intercept and slope model

- more general model allows for some form of nonstationarity—the variability within the subject now depends on the time
- the random effects are now $\boldsymbol{w}_i \sim N_2(\mathbf{0}, \mathbb{G})$ and for simplicity we assume that $\mathbb{G} = \nu^2\mathbb{I}$, the covariates are $\boldsymbol{z}_{ij} = (1, t_{ij})^\top$
- the variance of ε_j is $\nu^2(1 + t_j^2) + \sigma^2 + \tau^2$ and the whole variance/covariance structure of ε is $Var\varepsilon = \mathbb{Z}\mathbb{G}\mathbb{Z}^\top + \sigma^2\mathbb{H} + \tau^2\mathbb{I}$, where the matrix $\mathbb{Z}\mathbb{G}\mathbb{Z}^\top$ is a $n \times n$ matrix with elements $t_j t_k$, for $j, k = 1, \ldots, n$

Random intercept and slope model

- more general model allows for some form of nonstationarity—the variability within the subject now depends on the time
- the random effects are now $\boldsymbol{w}_i \sim N_2(\mathbf{0}, \mathbb{G})$ and for simplicity we assume that $\mathbb{G} = \nu^2\mathbb{I}$, the covariates are $\boldsymbol{z}_{ij} = (1, t_{ij})^\top$
- the variance of ε_j is $\nu^2(1 + t_j^2) + \sigma^2 + \tau^2$ and the whole variance/covariance structure of ε is $Var\varepsilon = \mathbb{Z}\mathbb{G}\mathbb{Z}^\top + \sigma^2\mathbb{H} + \tau^2\mathbb{I}$, where the matrix $\mathbb{Z}\mathbb{G}\mathbb{Z}^\top$ is a $n \times n$ matrix with elements $t_j t_k$, for $j, k = 1, \ldots, n$
- how can by this structure revealed with the sample variogram?

$$\widehat{\gamma}_i(u) = \frac{1}{2(n-u)} \sum_{j=u+1}^{n} \Big[\varepsilon_{ij} - \varepsilon_{i(j-u)}\Big]^2, \quad \text{for } u \in \{0, \ldots, n-1\}$$

Random effects and measurement error

- serial correlation appear to be quite natural feature for the longitudinal data analysis but in some applications the effects of the serial correlation may be dominated by the random effects and measurement errors
- particularly, if $\sigma^2 > 0$ is small compared to $\nu^2 + \tau^2$
- the model error ε_{ij} reduces to $\varepsilon \boldsymbol{z}_{ij}^\top \boldsymbol{w}_i + \omega_{ij}$ and the corresponding variance/covariance structure is of the form $Var \varepsilon = \mathbb{Z}\mathbb{G}\mathbb{Z}^\top + \tau^2 \mathbb{I}$
- the simplest scenario involves the random intercept only, meaning that $\mathbb{Z}\mathbb{G}\mathbb{Z}$ reduces to the matrix $\mathbb{J}$ and $Var \varepsilon = \nu^2 \mathbb{J} + \tau^2 \mathbb{I}$
- also called a “split-plot” model (because it is equivalent with the correlation induced by the randomization for a classical split-plot experiment)

Model selection & model building

- **Practical utilization of the model**
  - firstly, it is important to be able to validly answer the question of interest
  - statistical inference—statistical tests and confidence intervals/regions
- **Conditional mean structure**
  - exploratory in terms of some visualization tools (plots, graphs, etc.)
  - modeling in terms of the model matrix $\mathbb{X}$
- **Designe of experiment**
  - many existing problems could be avoided by a proper experiment planning
  - balanced data, proper randomization, treatment assignments, etc.
- **Variace/covariance structure**
  - exploratory in terms of some residuals inspection
  - effects of unobserved covariates,

Different covariance structures in SAS

See, for instance, the implementation of `PROC MIXED` in SAS and the corresponding SAS help/tutorial

- variance components
- compound symmetry
- unstructured
- autoregressive
- spatial
- ...

SAS Documentation at https://documentation.sas.com

Model diagnostics

The main idea of the statistical modeling process:

model formulation $\rightarrow$ **model estimation** $\rightarrow$ **statistical inference** $\rightarrow$ **model diagnostics**

Model diagnostics

The main idea of the statistical modeling process:

model formulation $\rightarrow$ **model estimation** $\rightarrow$ **statistical inference** $\rightarrow$ **model diagnostics**

- **The mean structure**
  - simple empirical characteristics, data scatterplots
  - simple summary plots/graphs (e.g., boxplots)
  - ...

- **The variance/covariance structure**
  - residual inspection and various residual plots
  - sample variogram function and its alternatives
  - empirical variogram vs. fitted variogram (model based)
  - ...

Summary

- exploratory analysis
- model building
- model diagnostics
- confirmatory analysis
- model interpretation