Lecture 8 | 14.04.2025

# Generalized linear models
## with random effects (GLMM)

# Generalization of LMMs

## Linear mixed effects model

- normal (or at least close to normal) data
- linear model + normality = "*lightness of being*"
- two basic modeling approaches (hierarchical vs. marginal model)
- relatively straightforward way from one model to another

# Generalization of LMMs

## Linear mixed effects model

- normal (or at least close to normal) data
- linear model + normality = "*lightness of being*"
- two basic modeling approaches (hierarchical vs. marginal model)
- relatively straightforward way from one model to another

## Generalized linear mixed effects model

- situations with non-normally distributed data
- introduction of some non-linearity in the model
- wide range of different modeling options (and different strategies)
- typically with no straightforward way to switch among the models

With nonlinear models, different assumptions imposed on the correlation structure of the data typically lead to regression coefficient estimates (the conditional mean structure in particular) that have different interpretation...

# Generalization of LMMs

## Linear mixed effects model

- normal (or at least close to normal) data
- linear model + normality = "*lightness of being*"
- two basic modeling approaches (hierarchical vs. marginal model)
- relatively straightforward way from one model to another

## Generalized linear mixed effects model

- situations with non-normally distributed data
- introduction of some non-linearity in the model
- wide range of different modeling options (and different strategies)
- typically with no straightforward way to switch among the models

With nonlinear models, different assumptions imposed on the correlation structure of the data typically lead to regression coefficient estimates (the conditional mean structure in particular) that have different interpretation...

**The objective of the analysis and the underlying sources of the variability/correlation within the data must assessed much more carefully ...**

# Typical non-normal data

In general, repeated (correlated) observations can be also measured with respect to some random variable $Y$ that is not necessarily continuous...
*(continuity nor even the Gaussian distribution can be assumed)*

Consider a longitudinal dataset $\{(\boldsymbol{Y}_i, \boldsymbol{X}_{i1}^\top, \dots, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N\}$ where $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ are repeated observations within the subject $i \in \{1, \dots, N\}$, for $N \in \mathbb{N}$ independent subjects with $n_i \in \mathbb{N}$ being the number of observations within the subject

- $Y_{ij} \in \mathbb{N} \cup \{0\}$ for all $i = 1, \dots, N$ and $j = 1, \dots, n_i$ (counts)
- $Y_{ij} \in \{1, \dots, K\}$ for all $i = 1, \dots, N$ and $j = 1, \dots, n_i$ (labels)
- $Y_{ij} \in \{0, 1\}$ for all $i = 1, \dots, N$ and $j = 1, \dots, n_i$ (true/false)
- ...

# Typical non-normal data

In general, repeated (correlated) observations can be also measured with respect to some random variable $Y$ that is not necessarily continuous...
*(continuity nor even the Gaussian distribution can be assumed)*

Consider a longitudinal dataset $\{(\boldsymbol{Y}_i, \boldsymbol{X}_{i1}^\top, \dots, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N\}$ where $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top$ are repeated observations within the subject $i \in \{1, \dots, N\}$, for $N \in \mathbb{N}$ independent subjects with $n_i \in \mathbb{N}$ being the number of observations within the subject

- $Y_{ij} \in \mathbb{N} \cup \{0\}$ for all $i = 1, \dots, N$ and $j = 1, \dots, n_i$ (counts)
- $Y_{ij} \in \{1, \dots, K\}$ for all $i = 1, \dots, N$ and $j = 1, \dots, n_i$ (labels)
- $Y_{ij} \in \{0, 1\}$ for all $i = 1, \dots, N$ and $j = 1, \dots, n_i$ (true/false)
- ...

In some sense, the non-normal data are even more frequent when analyzing the longitudinal studies (as it is always possible to transform some continuous response $Y \in \mathbb{R}$ into a categorical/binary information only... but not vise versa)

# GLM extensions for the longitudinal data

- **Marginal models**
  - primary (the solely) interest is given to the conditional mean structure
  - the correlation structure of the data is taken into account for inference
  - model interpretation with respect to the subpopulation comparisons

# GLM extensions for the longitudinal data

- **Marginal models**
  - primary (the solely) interest is given to the conditional mean structure
  - the correlation structure of the data is taken into account for inference
  - model interpretation with respect to the subpopulation comparisons

- **Random effects models**
  - substantial (theoretical/computational) issues due to nonlinearity
  - the interest may be given to a subject specific interpretation as well
  - the subject specific interpretation is, however, not straightforward...

# GLM extensions for the longitudinal data

- **Marginal models**
  - primary (the solely) interest is given to the conditional mean structure
  - the correlation structure of the data is taken into account for inference
  - model interpretation with respect to the subpopulation comparisons
- **Random effects models**
  - substantial (theoretical/computational) issues due to nonlinearity
  - the interest may be given to a subject specific interpretation as well
  - the subject specific interpretation is, however, not straightforward...
- **Transition models**
  - primary interest again with respect to the overall mean structure
  - the correlation structure is induced by the historical observations
  - wide range of different model structures and modeling techniques...

# GLM extensions for the longitudinal data

- **Marginal models**
  - primary (the solely) interest is given to the conditional mean structure
  - the correlation structure of the data is taken into account for inference
  - model interpretation with respect to the subpopulation comparisons

- **Random effects models**
  - substantial (theoretical/computational) issues due to nonlinearity
  - the interest may be given to a subject specific interpretation as well
  - the subject specific interpretation is, however, not straightforward...

- **Transition models**
  - primary interest again with respect to the overall mean structure
  - the correlation structure is induced by the historical observations
  - wide range of different model structures and modeling techniques...

↪ and possibly many others...

# 1. Marginal models

The conditional expectation of the response is modeled (using the set of explanatory variables) separately from the within-subject correlation. The conditional expectation is modeled by averaging the subjects which share the same values of the explanatory variables (basically the same is done in a standard linear regression model)

- **Mean structure**
  The marginal (conditional) expectation of the response depends (non-linearly) on a linear combination of the explanatory variables
  $$h(\mu_{ij}) = \boldsymbol{X}_{ij}^\top \beta, \qquad \text{for } \mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}] \text{ and } \boldsymbol{\beta} \in \mathbb{R}^p$$

- **Variance structure**
  The marginal (conditional) variance of the response depends on the marginal mean (and, optionally, some other parameters) as
  $$Var(Y_{ij}|\boldsymbol{X}_{ij}) = v(\mu_{ij})\phi, \quad \text{for } \phi > 0$$

- **Covariance structure**
  The correlation between two observations $Y_{ij}$ and $Y_{ik}$ (within the same subject $i \in \{1, \dots, N\}$) is assumed to be modeled as
  $$Cor(Y_{ij}, Y_{ik}|\boldsymbol{X}_{ij}, \boldsymbol{X}_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \boldsymbol{\alpha}), \quad \text{for } \boldsymbol{\alpha} \in \mathbb{R}^q$$

# Model interpretation (pros and cons)

- GLM models for the correlated/longitudinal data are natural analogues of classical GLM for independent observations

- Unknown parameters $\beta \in \mathbb{R}^p$ have the same (marginal) interpretation as the coefficients in a standard cross-sectional analysis
(i.e., the GLM regression model for independent observations)

- Depending on the domain of the (random) response variable $Y$, different models (with different interpretation of $\beta \in \mathbb{R}^p$) can be formulated
(logistic model, Poisson model, gamma model, etc.)

- All models so far (multivariate normal, marginal, hierarchical) can be (in some sense) interpreted as cross-sectional models (as all these models contain the term $\boldsymbol{X}^\top\beta$ somehow related to the conditional mean $E[Y|\boldsymbol{X}]$)

# Example: Logistic marginal model

The variable if interest, $Y$, takes only two possible values—some property is either achieved or, alternatively, it is not... The probability of achieving the given property, e.g. $P[Y = 1]$, is modeled conditionally on $\boldsymbol{X} \in \mathbb{R}^p$

For longitudinal data $\{(\boldsymbol{Y}_i^\top, \boldsymbol{X}_{i1}^\top, \dots, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N; j = 1, \dots, n_i\}$, where $\boldsymbol{Y}_i = (Y_{i1}, \dots . Y_{in_i})^\top$ represents correlated observations within the given subject $i \in \{1, \dots, N\}$ we primarily model $\mu_{ij} = E[Y_{ij} | \boldsymbol{X}_{ij}]$

# Example: Logistic marginal model

The variable if interest, $Y$, takes only two possible values—some property is either achieved or, alternatively, it is not... The probability of achieving the given property, e.g. $P[Y = 1]$, is modeled conditionally on $\boldsymbol{X} \in \mathbb{R}^p$

For longitudinal data $\{(\boldsymbol{Y}_i^\top, \boldsymbol{X}_{i1}^\top, \dots, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N; j = 1, \dots, n_i\}$, where $\boldsymbol{Y}_i = (Y_{i1}, \dots . Y_{in_i})^\top$ represents correlated observations within the given subject $i \in \{1, \dots, N\}$ we primarily model $\mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}]$

## One possible formulation

- $logit(\mu_{ij}) = \log \frac{\mu_{ij}}{1-\mu_{ij}} = \log \frac{P[Y_{ij}|\boldsymbol{X}_{ij}]}{1-P[Y_{ij}=1|\boldsymbol{X}_{ij}]} = \boldsymbol{X}_{ij}^\top \beta$
- $Var(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}(1 - \mu_{ij})$
- $Cor(Y_{ij}, Y_{ik}) = \alpha$, for any $i \in \{1, \dots, N\}$ and $j \neq k$

# Example: Logistic marginal model

The variable if interest, $Y$, takes only two possible values—some property is either achieved or, alternatively, it is not... The probability of achieving the given property, e.g. $P[Y = 1]$, is modeled conditionally on $\boldsymbol{X} \in \mathbb{R}^p$

For longitudinal data $\{(\boldsymbol{Y}_i^\top, \boldsymbol{X}_{i1}^\top, \dots, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N; j = 1, \dots, n_i\}$, where $\boldsymbol{Y}_i = (Y_{i1}, \dots . Y_{in_i})^\top$ represents correlated observations within the given subject $i \in \{1, \dots, N\}$ we primarily model $\mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}]$

## One possible formulation

- $logit(\mu_{ij}) = \log \frac{\mu_{ij}}{1-\mu_{ij}} = \log \frac{P[Y_{ij}|\boldsymbol{X}_{ij}]}{1-P[Y_{ij}=1|\boldsymbol{X}_{ij}]} = \boldsymbol{X}_{ij}^\top \boldsymbol{\beta}$
- $Var(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}(1 - \mu_{ij})$
- $Cor(Y_{ij}, Y_{ik}) = \alpha$, for any $i \in \{1, \dots, N\}$ and $j \neq k$

The estimated parameters $\widehat{\beta} = (\widehat{\beta}_1, \dots, \widehat{\beta}_p)^\top$ are interpreted in terms of the estimated odds ratios for true/false (i.e., $\exp\{\widehat{\beta}_j\}$, for $j \in \{1, \dots, p\}$)

# Correlation in the logistic marginal model

- Recall, that $Cor(Y_{ij}, Y_{ik}) = \alpha$ no matter what are the times $t_j$, or $t_k$ of the observations or the outcomes $\mu_{ij}$ and $\mu_{ik}$ (e.g., a random intercept model)
- For a binary outcomes $Y_1, Y_2 \in \{0,1\}$ with the means $\mu_1, \mu_2 \in (0,1)$ the correlation between $Y_1$ and $Y_2$ equals

$$Cor(Y_1, Y_2) = \frac{P[Y_1 = 1, Y_2 = 1] - \mu_1\mu_2}{\sqrt{(\mu_1(1-\mu_1)\mu_2(1-\mu_2)}}$$

# Correlation in the logistic marginal model

- Recall, that $Cor(Y_{ij}, Y_{ik}) = \alpha$ no matter what are the times $t_j$, or $t_k$ of the observations or the outcomes $\mu_{ij}$ and $\mu_{ik}$ (e.g., a random intercept model)
- For a binary outcomes $Y_1, Y_2 \in \{0, 1\}$ with the means $\mu_1, \mu_2 \in (0, 1)$ the correlation between $Y_1$ and $Y_2$ equals

$$Cor(Y_1, Y_2) = \frac{P[Y_1 = 1, Y_2 = 1] - \mu_1\mu_2}{\sqrt{(\mu_1(1-\mu_1)\mu_2(1-\mu_2)}}$$

- The joint probability $P[Y_1 = 1, Y_2 = 1]$ can be constrained as

$$\max\{0, \mu_1 + \mu_2 - 1\} < P[Y_1 = 1, Y_2 = 1] < \min\{\mu_1, \mu_2\}$$

and, therefore, the correlation $Cor(Y_1, Y_2)$ is constrained by a complicated form that depends on the marginal means $\mu_1$ and $\mu_2$

# Correlation in the logistic marginal model

- Recall, that $Cor(Y_{ij}, Y_{ik}) = \alpha$ no matter what are the times $t_j$, or $t_k$ of the observations or the outcomes $\mu_{ij}$ and $\mu_{ik}$ (e.g., a random intercept model)
- For a binary outcomes $Y_1, Y_2 \in \{0, 1\}$ with the means $\mu_1, \mu_2 \in (0, 1)$ the correlation between $Y_1$ and $Y_2$ equals

$$Cor(Y_1, Y_2) = \frac{P[Y_1 = 1, Y_2 = 1] - \mu_1\mu_2}{\sqrt{(\mu_1(1 - \mu_1)\mu_2(1 - \mu_2)}}$$

- The joint probability $P[Y_1 = 1, Y_2 = 1]$ can be constrained as

$$\max\{0, \mu_1 + \mu_2 - 1\} < P[Y_1 = 1, Y_2 = 1] < \min\{\mu_1, \mu_2\}$$

  and, therefore, the correlation $Cor(Y_1, Y_2)$ is constrained by a complicated form that depends on the marginal means $\mu_1$ and $\mu_2$
- The odds ratio are used instead to model the association between binary observations

$$OR(Y_1, Y_2) = \frac{P[Y_1 = 1, Y_2 = 1]P[Y_1 = 0, Y_2 = 0]}{P[Y_1 = 1, Y_2 = 0]P[Y_1 = 0, Y_2 = 1]}$$

  which is not constrained by the means $\mu_1$ and $\mu_2$ any more (rather than the correlation, the odds ratio are equal to some constant, $\alpha \in \mathbb{R}$)

# Example: Logarithmic model for Poisson counts

The variable if interest, $Y \in \mathbb{N} \cup \{0\}$, takes infinitely many (ordinal) values—counts—and the main objective is to model the conditional mean $\mu = E[Y|\boldsymbol{X}] \in \mathbb{R}_+$ where, again $\boldsymbol{X} \in \mathbb{R}^p$

For longitudinal data $\{(\boldsymbol{Y}_i^\top, \dots, \boldsymbol{X}_{i1}^\top, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N; j = 1, \dots, n_i\}$, where $\boldsymbol{Y}_i = (Y_{i1}, \dots . Y_{in_i})^\top$ represents correlated observations within the given subject $i \in \{1, \dots, N\}$ we again primarily model $\mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}]$

## One possible formulation

- $log(\mu_{ij}) = \boldsymbol{X}_{ij}^\top \beta$
- $Var(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}\phi_{ij}$
- $Cor(Y_{ij}, Y_{ik}) = \alpha$

# Example: Logarithmic model for Poisson counts

The variable if interest, $Y \in \mathbb{N} \cup \{0\}$, takes infinitely many (ordinal) values—counts—and the main objective is to model the conditional mean $\mu = E[Y|\boldsymbol{X}] \in \mathbb{R}_+$ where, again $\boldsymbol{X} \in \mathbb{R}^p$

For longitudinal data $\{(\boldsymbol{Y}_i^\top, \dots, \boldsymbol{X}_{i1}^\top, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N; j = 1, \dots, n_i\}$, where $\boldsymbol{Y}_i = (Y_{i1}, \dots . Y_{in_i})^\top$ represents correlated observations within the given subject $i \in \{1, \dots, N\}$ we again primarily model $\mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}]$

## One possible formulation

- $log(\mu_{ij}) = \boldsymbol{X}_{ij}^\top \beta$
- $Var(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}\phi_{ij}$
- $Cor(Y_{ij}, Y_{ik}) = \alpha$

The correlation within the repeated observations is sometimes ignored, or it is modeled simply, or models with random effects are more appropriate

# 2. Random effects models

A straightforward extension of the linear mixed models (LMMs) for discrete and non-Gaussian responses where it is assumed that the response is independent for subjects and follows a GLM model but the regression (mean) coefficients may vary from subject to subject (with correlated observations within each subject)

- **Mean structure**

$$\mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i] = \psi'(\theta_{ij})$$

- **Variance structure**

$$v_{ij} = Var[Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i] = \psi''(\theta_{ij})\phi$$

where we assume the exponential family for the conditional distribution of $Y_{ij}|(\boldsymbol{X}_{ij}, \boldsymbol{w}_i)$ with $f_{(Y|\boldsymbol{X}_{ij},\boldsymbol{w}_i)}(y) = \exp\{[y\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y,\phi)\}$, where $g(\mu_{ij}) = \boldsymbol{X}_{ij}^\top\beta + \boldsymbol{Z}_{ij}^\top\boldsymbol{w}_i$ and $v_{ij} = v(\mu_{ij})\phi$ (link and variance functions)

- **Covariance structure**
  Random effects $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N$ are independent with some common underlying distribution and the subject specific responses $Y_{i1}, \ldots, Y_{in_i}$ are, conditionally on $\boldsymbol{w}_i$, independent

# Some stochastic properties

- Models with random effects assume that there is some natural heterogeneity across individuals (reflected in their regression coefficients) and this heterogeneity can be represented by some probability distribution

- The correlation structure of the repeated observations within each subject is modeled more carefully than in the marginal models but due to non-linearity of the model, the interpretation is more challenging

- The correlation within the repeated observations for one subject arises from sharing the same random effect $\boldsymbol{w}_i$ where the random effect is actually not observed (so-called latent variable models)

- The random effects models are primarily used when the main objective is to perform some statistical inference about individuals rather than the population (sub-populations) means only

# Example: Logistic regression with random effects

The variable of interest, $Y$, takes only two possible values—e.g., some property is either achieved or it is not... The probability of achieving the given property, e.g. $P[Y = 1]$, is modeled conditionally on $\boldsymbol{X} \in \mathbb{R}^p$

For the longitudinal data $\{(\boldsymbol{Y}_i^\top, \boldsymbol{X}_{i1}^\top, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N; j = 1, \dots, n_i\}$, where $\boldsymbol{Y}_i = (Y_{i1}, \dots . Y_{in_i})^\top$ represents correlated observations within the subject $i \in \{1, \dots, N\}$ we model $\mu_{ij} = E[Y_{ij} | \boldsymbol{X}_{ij}, \boldsymbol{w}_i]$

## One possible formulation

- $logit(\mu_{ij}) = \log \frac{\mu_{ij}}{1-\mu_{ij}} = \log \frac{P[Y_{ij} | \boldsymbol{X}_{ij}, \boldsymbol{w}_i]}{1 - P[Y_{ij} = 1 | \boldsymbol{X}_{ij}, \boldsymbol{w}_i]} = \boldsymbol{X}_{ij}^\top \beta + \boldsymbol{Z}_{ij}^\top \boldsymbol{w}_i$
- $Var(Y_{ij} | \boldsymbol{X}_{ij}, \boldsymbol{w}_i) = \mu_{ij}(1 - \mu_{ij})$
- typically $\boldsymbol{w}_i \sim N(\boldsymbol{0}, \mathbb{G})$ and repeated observations within the subject are, given $\boldsymbol{w}_i$ independent

# Example: Logistic regression with random effects

The variable of interest, $Y$, takes only two possible values—e.g., some property is either achieved or it is not... The probability of achieving the given property, e.g. $P[Y=1]$, is modeled conditionally on $\boldsymbol{X} \in \mathbb{R}^p$

For the longitudinal data $\{(\boldsymbol{Y}_i^\top, \boldsymbol{X}_{i1}^\top, \boldsymbol{X}_{in_i}^\top)^\top;\ i = 1, \dots, N; j = 1, \dots, n_i\}$, where $\boldsymbol{Y}_i = (Y_{i1}, \dots . Y_{in_i})^\top$ represents correlated observations within the subject $i \in \{1, \dots, N\}$ we model $\mu_{ij} = E[Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i]$

## One possible formulation

- $logit(\mu_{ij}) = \log \frac{\mu_{ij}}{1-\mu_{ij}} = \log \frac{P[Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i]}{1-P[Y_{ij}=1|\boldsymbol{X}_{ij}, \boldsymbol{w}_i]} = \boldsymbol{X}_{ij}^\top \beta + \boldsymbol{Z}_{ij}^\top \boldsymbol{w}_i$
- $Var(Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{w}_i) = \mu_{ij}(1-\mu_{ij})$
- typically $\boldsymbol{w}_i \sim N(\boldsymbol{0}, \mathbb{G})$ and repeated observations within the subject are, given $\boldsymbol{w}_i$ independent

The random effects model is most useful when the main objective is to make inference about individuals, rather than the populations averages

# 3. Transition models

Unlike the previous models, the correlation among the repeated observations within a given subjects, $Y_{i1}, \dots, Y_{in_i}$, occurs because there is an explicit influence of the previous values $\boldsymbol{Y}_i^{(j)} = (Y_{i1}, \dots, Y_{i(j-1)})^\top$ on the most recent one $Y_{ij}$ and the historical observations are treated as additional regresors in the model

- **Typical mean structure**

$$\mu_{ij} = E[Y_{ij} | \boldsymbol{X}_{ij}, Y_{i(j-1)}, \dots, Y_{i1}] = \boldsymbol{X}_{ij}^\top \beta + \boldsymbol{Y}_i^{(j)\top} \gamma$$

- **Variance structure**

$$Var[Y_{ij} | \boldsymbol{X}_{ij}, Y_{i(j-1)}, \dots, Y_{i1}] = \psi''(\theta_{ij})\phi$$

where $f_{(Y_j | \boldsymbol{X}_{ij}, \boldsymbol{Y}_i^{(j)})}(y) = \exp\{[y\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y, \phi)\}$, where $g(\mu_{ij}) = \boldsymbol{X}_{ij}^\top \beta + \boldsymbol{Y}_i^{(j)\top} \gamma$

- **Covariance structure**
  Directly induced by the previous (historical) observations on the given subject $i \in \{1, \dots, N\}$

# Some stochastic properties

- A simple marginal model with an exponential autocorrelation function $Cor(Y_{ij}, Y_{ik}) = \sigma^2 \exp\{-\phi|t_j - t_k|\}$ can be also formulated as a transition model where

$$Y_{ij} = \boldsymbol{X}_{ij}^\top \beta + \varepsilon_{ij}, \qquad \text{where } \varepsilon_{ij} = \alpha\varepsilon_{i(j-1)} + \omega_{ij},$$

with $\alpha = \exp\{-\phi\}$ and $\omega_{ij} \sim N(0, \tau^2)$, for $\tau^2 = \sigma^2(1 - \alpha^2)$

- By substituting $\varepsilon_{ij} = Y_{ij} - \boldsymbol{X}_{ij}^\top \beta$ into $\varepsilon_{ij} = \alpha\varepsilon_{i(j-1)} + \omega_{ij}$ we have

$$Y_{ij}|\boldsymbol{X}_{ij}, Y_{i(j-1)} \sim N(\boldsymbol{X}_{ij}^\top \beta + \alpha(Y_{i(j-1)} - \boldsymbol{X}_{i(j-1)}^\top \beta), \tau^2)$$

which can be easily specified (typically for situations where many repeated observations within subjects are available)

# Theoretical and empirical comparisons

In a simple linear case it is possible to use three regression approaches in a way that the models have the coefficients with the same interpretation

- Consider a simple linear regression model $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij}$ where $E[Y_{ij}|t_{ij}] = \beta_0 + \beta_1 t_{ij}$ and $Cor(Y_{ij}, Y_{ik}) = \rho(t_{ij}, t_{ik}, \boldsymbol{\alpha})$
- Consider a two-stage correlation structure, where $\rho(t_{ij}, t_{ik}, \boldsymbol{\alpha}) = \alpha_0$ for $|t_{ij} - t_{ik}| < t_0$ and $\rho(t_{ij}, t_{ik}, \boldsymbol{\alpha}) = \alpha_1$ for $|t_{ij} - t_{ik}| \geq t_0$

# Theoretical and empirical comparisons

In a simple linear case it is possible to use three regression approaches in a way that the models have the coefficients with the same interpretation

- Consider a simple linear regression model $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij}$ where $E[Y_{ij}|t_{ij}] = \beta_0 + \beta_1 t_{ij}$ and $Cor(Y_{ij}, Y_{ik}) = \rho(t_{ij}, t_{ik}, \boldsymbol{\alpha})$
- Consider a two-stage correlation structure, where $\rho(t_{ij}, t_{ik}, \boldsymbol{\alpha}) = \alpha_0$ for $|t_{ij} - t_{ik}| < t_0$ and $\rho(t_{ij}, t_{ik}, \boldsymbol{\alpha}) = \alpha_1$ for $|t_{ij} - t_{ik}| \geq t_0$

- **Marginal model** approaches the mean and the correlation structure separately (straightforward interpretation)
- **Mixed model** can be written as $Y_{ij} = \beta_0 + w_{i0} + (\beta_1 + w_{i1}) t_{ij} + \omega_{ij}$ , where $\omega_{ij} \sim N(0, \sigma^2)$ (independent) and $\boldsymbol{w}_i = (w_{i0}, w_{i1})^\top \sim N(\boldsymbol{0}, \mathbb{G})$
- **Transition model** can be formulated as $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \varepsilon_{ij}$, where $\varepsilon_{ij} = \alpha \varepsilon_{i(j-1)} + \omega_{ij}$, with $\omega_{ij} \sim N(0, \sigma^2)$ (independent)

# Summary

- For non-normal data there is a wide class of different modeling approaches *(depending on the distribution of the response—however, within the exponential family)*
- The most common regression models (among others) include
  - marginal models
  - random-effects models
  - transition models
- Different models are based on different sets of assumptions and, therefore, different interpretation options follow from each method
- Some models are more suitable for modeling the population mean structure only, others are more suitable when the inference is about to be performed with respect to individual subjects
- Each class of these models will be discussed in more details later...