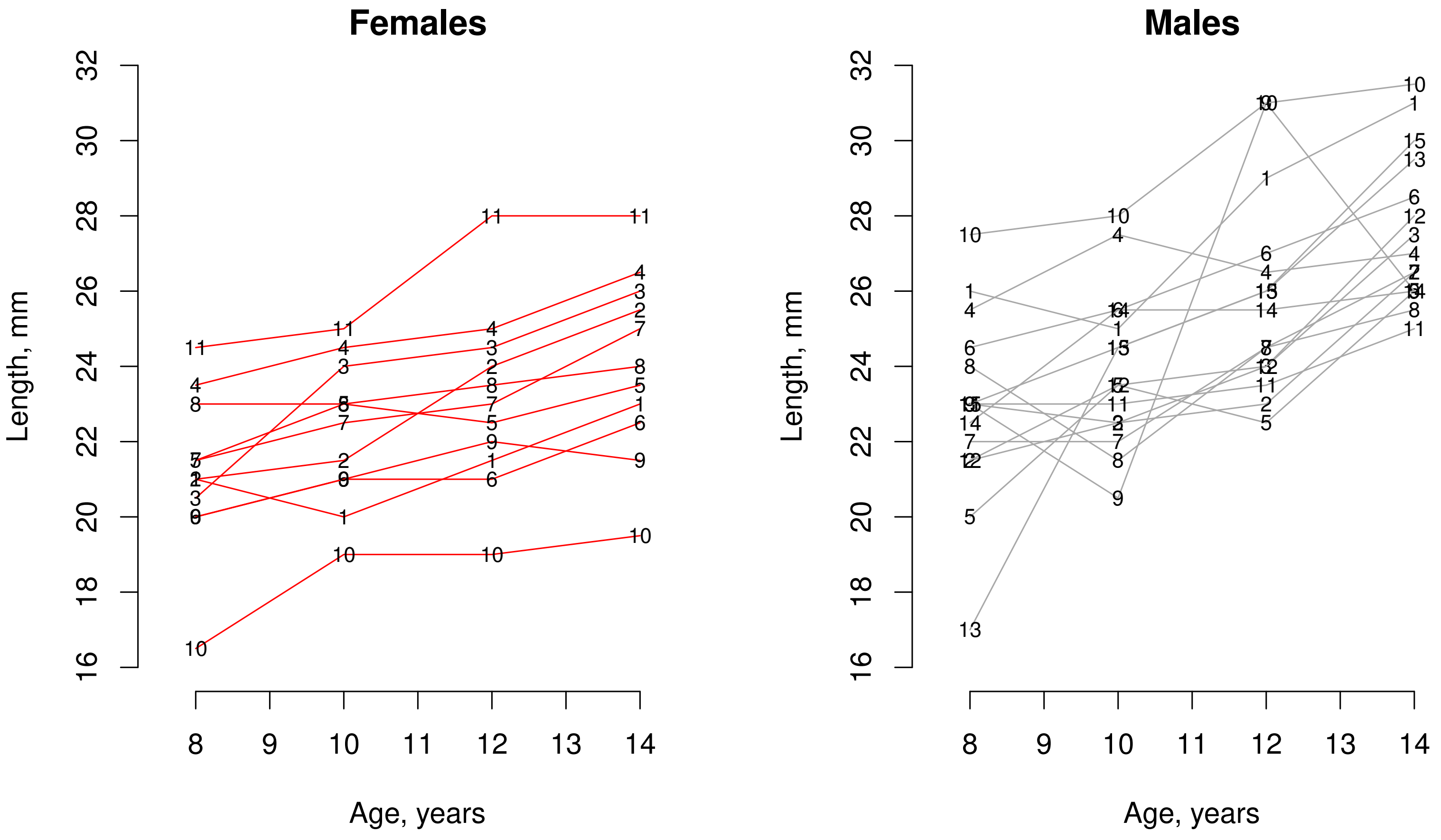
Part III:

Linear mixed effects models

Dental growth data



- One possible scientific goal is to estimate and formally compare the average growth trajectory between males and females

- In Part II we considered *two-stage least squares* as one possible analysis strategy:

(1) estimate the growth trajectory for each child:

$$\boldsymbol{Y}_k \;=\; \boldsymbol{Z}_k\boldsymbol{\beta}_k \;+\; \boldsymbol{\epsilon}_k$$

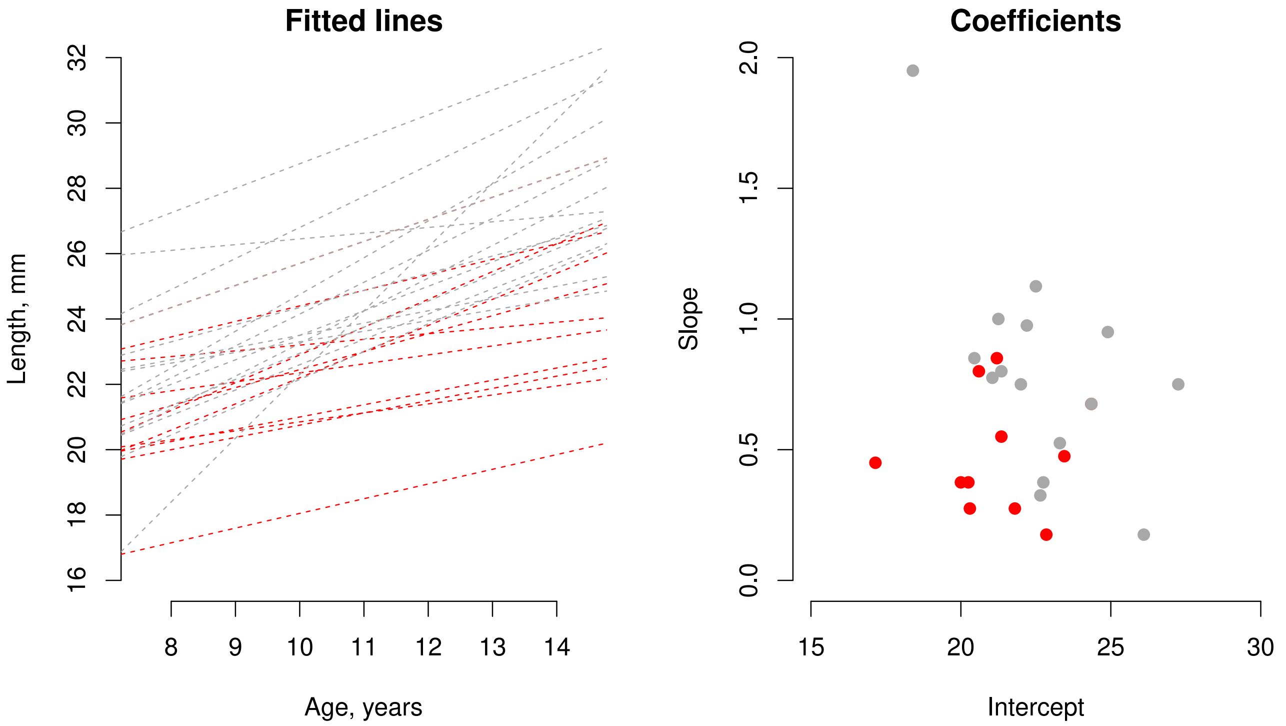
- ∗ child-specific regression parameters, $\boldsymbol{\beta}_k$
- ∗ $\boldsymbol{\epsilon}_k$ are observation-specific random variations around each subjects' underlying growth curve
- ∗ assumed to be i.i.d with mean 0 and variance σ_k^2

(2) consider differences in the child-specific coefficients between males and females

$$\boldsymbol{\beta}_k \;=\; \boldsymbol{W}_k\boldsymbol{\beta} \;+\; \boldsymbol{\gamma}_k$$

- ∗ $\boldsymbol{\beta}$ characterizes the mean growth curve in the population (of children)
- ∗ $\boldsymbol{\gamma}_k$ are child-specific deviations around the population growth curve
- ∗ assumed to be i.i.d with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{G}$

- Results:



```
> summary(lm(beta1 ~ gender, data=betaMat))
...
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.1591     0.2280   0.698    0.492
gender        0.3205     0.1380   2.322    0.029 *
...
```

Issues

- The design matrix is constrained at each stage of the analysis
  - ⋆ stage 1 is restricted to within-subject covariates
  - ⋆ stage 2 is restricted to between-subject covariates
- Information is lost by having summarized the response vector for subject k at stage 1
- Differential uncertainty in the $\widehat{\boldsymbol{\beta}}_k$ is not accounted for at stage 2
  - ⋆ estimates, not the 'true' $\boldsymbol{\beta}_k$
  - ⋆ uncertainty depends on n_k and when the observations are collected
- The fact that observations are correlated is ignored
- All of these are key motivators for combining stages 1 and 2 into a single model formulation

Linear mixed effects models

Basic idea

- Assume that each cluster has a regression model characterized by cluster-specific regression parameters
- Structure the cluster-specific regressions across the population of clusters via a series of
    - ⋆ p *fixed effects* parameters that are common to all clusters in the population
    - ⋆ q *random effects* parameters that permit cluster-specific perturbations
- As we will see, both sets of parameters allow for differences in the regressions across subjects

Notation

$\boldsymbol{Y}_k = (Y_{k1},\ Y_{k2},\ \ldots,\ Y_{kn_k})^T$ Response vector

$\boldsymbol{\beta} = (\beta_1,\ \beta_2,\ \ldots,\ \beta_p)^T$ Fixed effects

$\boldsymbol{X}_{ki} = (X_{ki,1},\ X_{ki,2},\ \ldots,\ X_{ki,p})$

$\boldsymbol{X}_k = (\boldsymbol{X}_{k1},\ \boldsymbol{X}_{k2},\ \ldots,\ \boldsymbol{X}_{kn_k})^T$ Design matrix for the fixed effects

- ⋆ $n_k \times p$

$\boldsymbol{\gamma}_k = (\gamma_{k1},\ \gamma_{k2},\ \ldots,\ \gamma_{kq})^T$ Random effects

$\boldsymbol{Z}_{ki} = (Z_{ki,1},\ Z_{ki,2},\ \ldots,\ Z_{ki,q})$

$\boldsymbol{Z}_k = (\boldsymbol{Z}_{k1},\ \boldsymbol{Z}_{k2},\ \ldots,\ \boldsymbol{Z}_{kn_k})^T$ Design matrix for the random effects

- ⋆ $n_k \times q$

$\boldsymbol{\epsilon}_k = (\epsilon_{k1},\ \epsilon_{k2},\ \ldots,\ \epsilon_{kn_k})^T$ Residual error terms

Definition

- By a *linear mixed effects model*, we mean a statistical model with the following assumptions/components:
  - ⋆ model for the response of the k^{th} cluster, given the fixed and random effects:

$$\boldsymbol{Y}_k = \boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{Z}_k\boldsymbol{\gamma}_k + \boldsymbol{\epsilon}_k$$

  - ⋆ model for the random effects and residual error:

$$\begin{aligned} \mathsf{E}[\boldsymbol{\gamma}_k] &= \mathbf{0} & \mathsf{Cov}[\boldsymbol{\gamma}_k] &= \boldsymbol{G}(\boldsymbol{\alpha}) \\ \mathsf{E}[\boldsymbol{\epsilon}_k] &= \mathbf{0} & \mathsf{Cov}[\boldsymbol{\epsilon}_k] &= \boldsymbol{R}_k(\boldsymbol{\alpha}) \\ \mathsf{Cov}[\boldsymbol{\gamma}_k, \boldsymbol{\epsilon}_k] &= \mathbf{0} \end{aligned}$$

    - ∗ $\boldsymbol{\alpha}$ referred to as *covariance parameters*
  - ⋆ responses across clusters are independent of eachother

- Note, this specification defines a marginal linear model:

$$\begin{aligned}
\mathsf{E}[\boldsymbol{Y}_k|\ \boldsymbol{X}_k] &= \mathsf{E}_{\gamma}[\mathsf{E}_{\boldsymbol{Y}}[\boldsymbol{Y}_k|\ \boldsymbol{X}_k,\ \boldsymbol{\gamma}_k]] \\
&= \mathsf{E}_{\gamma}[\boldsymbol{X}_k\boldsymbol{\beta}\ +\ \boldsymbol{Z}_k\boldsymbol{\gamma}_k] \\
&= \boldsymbol{X}_k\boldsymbol{\beta} \\
&\equiv \boldsymbol{\mu}_k(\boldsymbol{\beta})
\end{aligned}$$

$$\begin{aligned}
\mathsf{Cov}[\boldsymbol{Y}_k|\ \boldsymbol{X}_k] &= \mathsf{Cov}_{\gamma}[\mathsf{E}_{\boldsymbol{Y}}[\boldsymbol{Y}_k|\ \boldsymbol{X}_k,\ \boldsymbol{\gamma}_k]]\ +\ \mathsf{E}_{\gamma}[\mathsf{Cov}_{\boldsymbol{Y}}[\boldsymbol{Y}_k|\ \boldsymbol{X}_k,\ \boldsymbol{\gamma}_k]] \\
&= \mathsf{Cov}_{\gamma}[\boldsymbol{X}_k\boldsymbol{\beta}\ +\ \boldsymbol{Z}_k\boldsymbol{\gamma}_k]\ +\ \mathsf{E}_{\gamma}[\boldsymbol{R}_k(\boldsymbol{\alpha})] \\
&= \boldsymbol{Z}_k\boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_k^T\ +\ \boldsymbol{R}_k(\boldsymbol{\alpha}) \\
&\equiv \boldsymbol{\Sigma}_k(\boldsymbol{\alpha})
\end{aligned}$$

$$\mathsf{Cov}[\boldsymbol{Y}_k, \boldsymbol{Y}_{k'}] = \boldsymbol{0}$$

- In principle, therefore, we could use the methods developed in Part II to estimate components of this model

- From the expression for $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$, we can see that $\boldsymbol{G}(\boldsymbol{\alpha})$ and $\boldsymbol{R}_k(\boldsymbol{\alpha})$ jointly characterize variation in the observed responses around the population mean induced by $\boldsymbol{\beta}$

$$\boldsymbol{\Sigma}_k(\boldsymbol{\alpha}) \;=\; \boldsymbol{Z}_k \boldsymbol{G}(\boldsymbol{\alpha}) \boldsymbol{Z}_k^T \;+\; \boldsymbol{R}_k(\boldsymbol{\alpha})$$

  - ⋆ $\boldsymbol{G}(\boldsymbol{\alpha})$ is a $q \times q$ matrix that quantifies random variation in the trajectories <u>across</u> clusters
  - ⋆ $\boldsymbol{R}_k(\boldsymbol{\alpha})$ is a $n_k \times n_k$ matrix that quantifies random variation <u>within</u> subjects

Interpretation

- Consider the interpretation of $\boldsymbol{\beta}$ in:

$$\boldsymbol{Y}_k \;=\; \boldsymbol{X}_k\boldsymbol{\beta} \;+\; \boldsymbol{Z}_k\boldsymbol{\gamma}_k \;+\; \boldsymbol{\epsilon}_k$$

- Following standard principles, the components of $\boldsymbol{\beta}$ correspond to differences in the mean response between specific populations
  - ⋆ determined by the specific covariate under consideration
  - ⋆ require 'holding other terms constant', including the random effects
  - ⋆ interpretation is therefore *conditional* on the random effect

- Since the γ_k are continuous, in theory they are unique to each cluster
  - ⋆ interpretation of $\boldsymbol{\beta}$ is therefore *cluster-specific*

- Some folks find this unappealing
  - ⋆ simultaneously thinking about averages and yet we are conditioning on a specific cluster

- Recall, however, the induced marginal model:

$$\mathsf{E}[\boldsymbol{Y}_k|\ \boldsymbol{X}_k] \ = \ \boldsymbol{X}_k\boldsymbol{\beta}$$

  - ⋆ the interpretation of $\boldsymbol{\beta}$ in this model does not require conditioning on the subject
  - ⋆ interpretation is therefore *marginal* with respect to cluster membership
  - ⋆ more in line with our 'usual' interpretation of regression parameters

- The main point here is that despite the model being conditionally specified, since the 'true' conditional and marginal parameters (i.e. the regression coefficients) are the same numerically one can interpret results from a linear mixed effects model as pertaining to marginal associations
  - ⋆ relevant key question becomes *how* do you estimate $\boldsymbol{\beta}$ as well as perform inference?

Q: What about the interpretation of the random effects?

  - ⋆ often interpreted as *latent* characteristics that are specific to the subject
  - ⋆ represent the collective impact of factors relevant to the response but not included in the model

- Need to be careful, however, not to interpret their inclusion as ‘adjusting’ for unmeasured confounding

Special case #1: Random intercepts model

- Arguably the simplest mixed effects model is one in which a single random intercept is introduced:

$$Y_{ki} \;=\; \boldsymbol{X}_{ki}^T\boldsymbol{\beta} \;+\; \gamma_{0k} \;+\; \epsilon_{ki}$$

  - ⋆ γ_{0k} is a cluster-specific deviation around the population intercept, β_0
  - ⋆ assume E$[\gamma_{0k}] = 0$ and V$[\gamma_{0k}] = \sigma^2_\gamma$
  - ⋆ assume E$[\epsilon_{ki}] = 0$ and V$[\epsilon_{ki}] = \sigma^2_\epsilon$

- Consider the induced marginal covariance, $\boldsymbol{\Sigma}_k$, by considering two specific study units:

$$\begin{aligned} Y_{ki} &= \boldsymbol{X}_{ki}^T\boldsymbol{\beta} \;+\; \gamma_{0k} \;+\; \epsilon_{ki} \\ Y_{kj} &= \boldsymbol{X}_{kj}^T\boldsymbol{\beta} \;+\; \gamma_{0k} \;+\; \epsilon_{kj} \end{aligned}$$

  - ⋆ notice how the two observations ‘share’ the same value of γ_{0k}

- We then have:

$$\begin{aligned}
\mathsf{V}[Y_{ki}] &= \mathsf{V}_\gamma[\mathsf{E}_Y[Y_{ki}|\ \gamma_{0k}]] \ + \ \mathsf{E}_\gamma[\mathsf{V}_Y[Y_{ki}|\ \gamma_{0k}]] \\
&= \sigma^2_\gamma \ + \ \sigma^2_\epsilon
\end{aligned}$$

$$\begin{aligned}
\mathsf{Cov}[Y_{ki}, Y_{kj}] &= \mathsf{Cov}_\gamma[\mathsf{E}_Y[Y_{ki}|\ \gamma_{0k}], \mathsf{E}_Y[Y_{kj}|\ \gamma_{0k}]] \\
&\qquad + \ \mathsf{E}_\gamma[\mathsf{Cov}_Y[Y_{ki}, Y_{kj}|\ \gamma_{0k}]] \\
&= \sigma^2_\gamma
\end{aligned}$$

so that for $\boldsymbol{\alpha} = (\sigma^2_\gamma,\ \sigma^2_\epsilon)$:

$$\boldsymbol{\Sigma}(\boldsymbol{\alpha}) \ = \ \begin{bmatrix} \sigma^2_\gamma + \sigma^2_\epsilon & \sigma^2_\gamma & \ldots & \sigma^2_\gamma \\ \sigma^2_\gamma & \sigma^2_\gamma + \sigma^2_\epsilon & \ldots & \sigma^2_\gamma \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_\gamma & \sigma^2_\gamma & \ldots & \sigma^2_\gamma + \sigma^2_\epsilon \end{bmatrix}$$

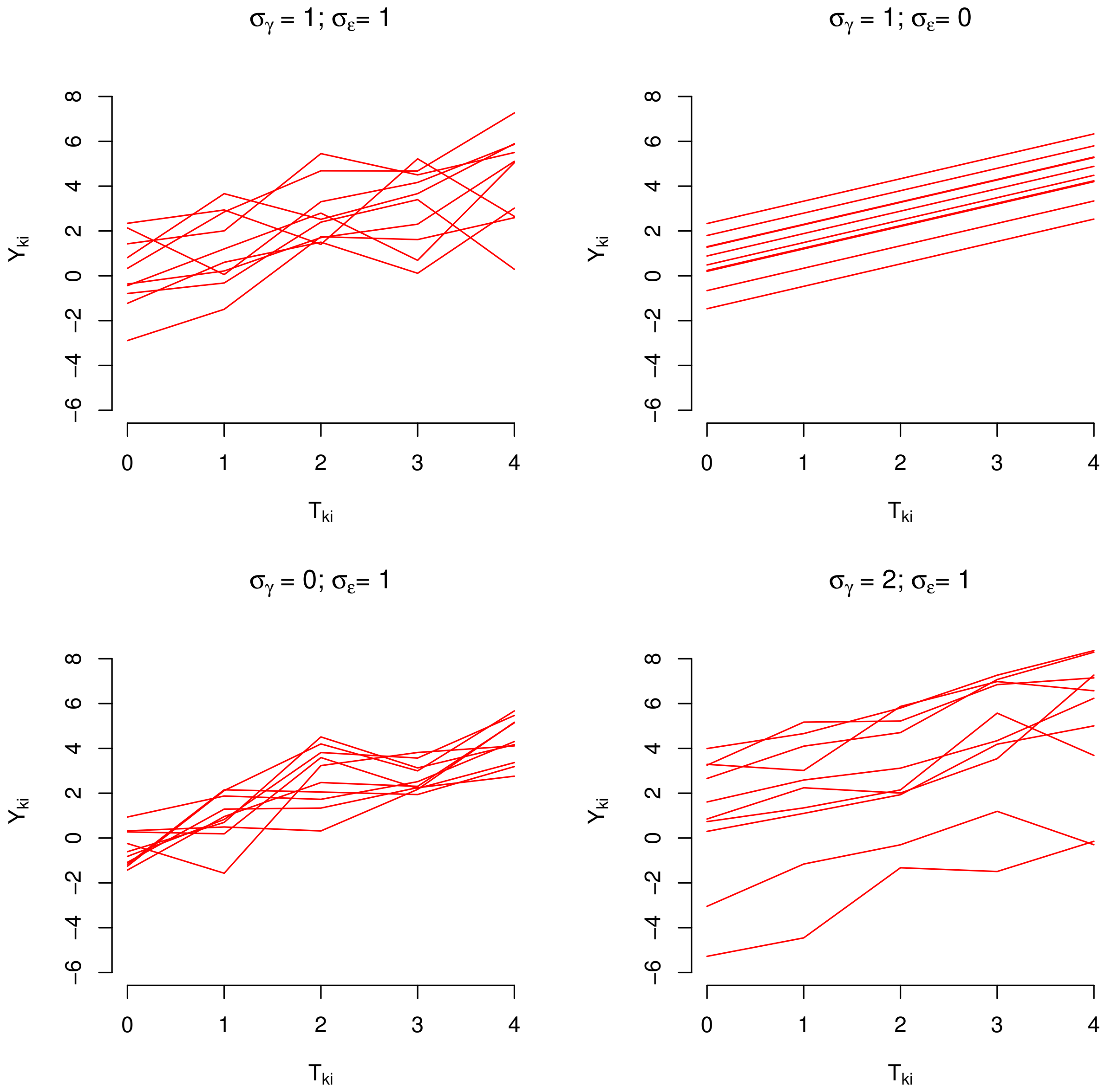
- Consequently, the induced marginal model has an exchangeable dependence structure

- Notice that the correlation between two study units is:

$$\text{Cor}[Y_{ki}, Y_{kj}] \;=\; \frac{\sigma_\gamma^2}{\sigma_\gamma^2 \;+\; \sigma_\epsilon^2} \;=\; \frac{\text{"Between"}}{\text{"Between"} \;+\; \text{"Within"}} \;\in [0, 1]$$

  - ⋆ only positive correlation is entertained
  - ⋆ degree of correlation depends on the interplay between the ‘between-cluster’ variation and the ‘within-cluster” variation in the responses
  - ⋆ if $\sigma_\gamma^2 = 0$ there is no correlation
  - ⋆ if $\sigma_\epsilon^2 = 0$ there is perfect (positive) correlation

- Visualization of $K\!=\!10$ clusters with $n_k\!=\!5$:



Special case #2: Random intercepts/slopes model

- The random intercept model can be extended to permit the effects of select covariates to vary across the clusters

- For example in a longitudinal study in which T_{ki} is the timing of the i^{th} observation in the k^{th} cluster, we might specify the model:

$$Y_{ki} \;=\; \boldsymbol{X}_{ki}^T\boldsymbol{\beta} \;+\; \gamma_{0k} \;+\; T_{ki}\gamma_{1k} \;+\; \epsilon_{ki}$$

  - ⋆ γ_{0k} is a cluster-specific deviation around the population intercept
  - ⋆ γ_{1k} is a cluster-specific deviation around the population slope for time
    - ∗ assume T_{ki} is an element of $\boldsymbol{X}_{ki}$
  - ⋆ assume

$$\mathsf{Cov}[\boldsymbol{\gamma}_k] \;=\; \boldsymbol{G}(\boldsymbol{\alpha}) \;=\; \begin{bmatrix} \Sigma_{\gamma,00} & \Sigma_{\gamma,01} \\ \Sigma_{\gamma,01} & \Sigma_{\gamma,11} \end{bmatrix}$$

  - ⋆ assume $\mathsf{E}[\epsilon_{ki}] = 0$ and $\mathsf{V}[\epsilon_{ki}] = \sigma_\epsilon^2$

- Again consider the induced marginal covariance, $\mathbf{\Sigma}_k$, by considering two specific study units:

$$\begin{aligned} Y_{ki} &= \beta_0 \ + \ \beta_1 T_{ki} \ + \ \gamma_{0k} \ + \ \gamma_{1k} T_{ki} \ + \ \epsilon_{ki} \\ Y_{kj} &= \beta_0 \ + \ \beta_1 T_{kj} \ + \ \gamma_{0k} \ + \ \gamma_{1k} T_{kj} \ + \ \epsilon_{kj} \end{aligned}$$

  - ⋆ for simplicity, just have the intercept and a term for T_{ki} in $\boldsymbol{X}_k$

- We then have:

$$\begin{aligned} \mathsf{V}[Y_{ki}] &= \mathsf{V}_{\boldsymbol{\gamma}}[\mathsf{E}_Y[Y_{ki}|\,\boldsymbol{\gamma}_k]] \ + \ \mathsf{E}_{\boldsymbol{\gamma}}[\mathsf{V}_Y[Y_{ki}|\,\boldsymbol{\gamma}_k]] \\ &= \Sigma_{\gamma,00} \ + \ 2\Sigma_{\gamma,01} T_{ki} \ + \ \Sigma_{\gamma,11} T_{ki}^2 \ + \ \sigma_\epsilon^2 \end{aligned}$$

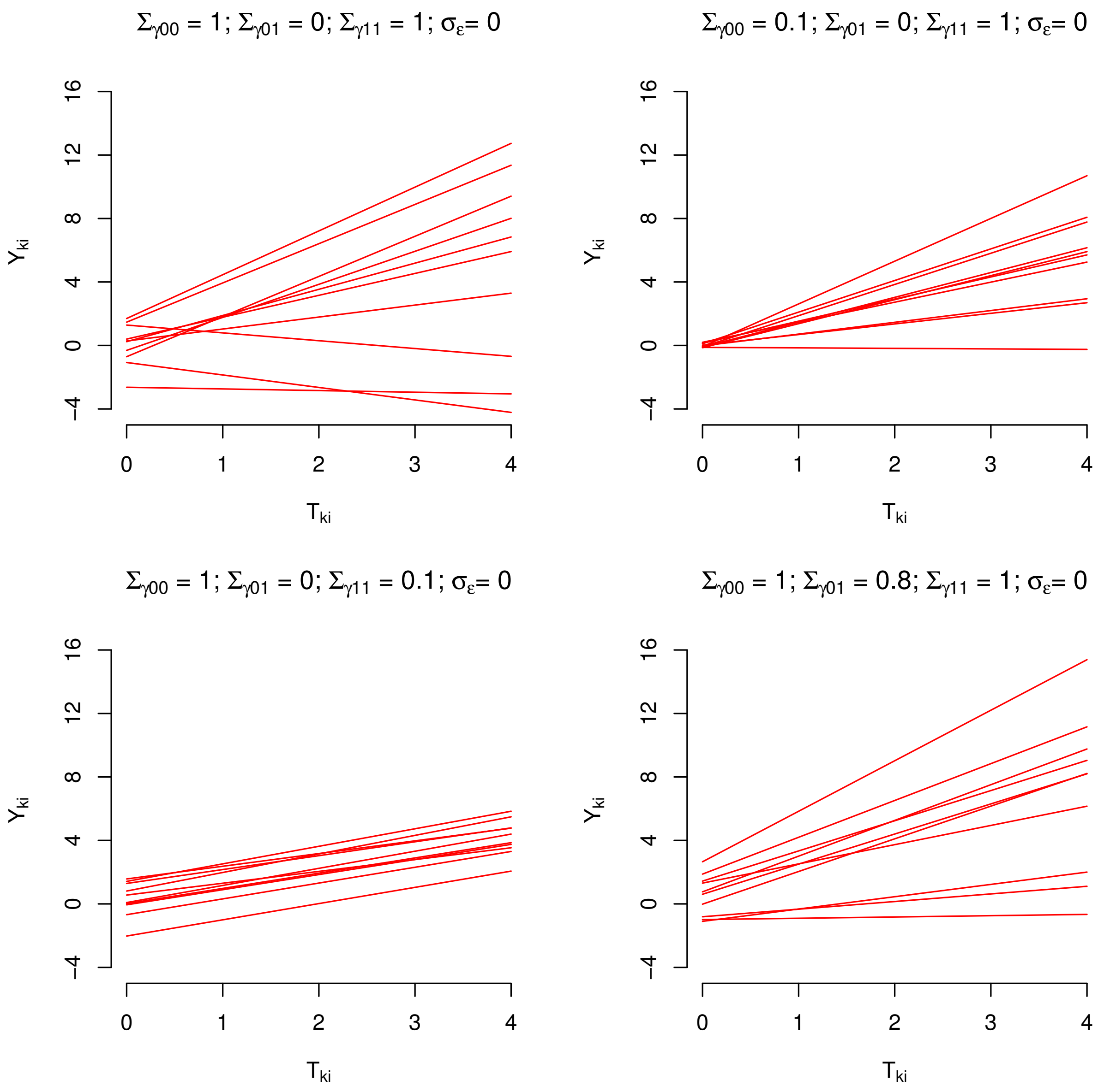
$$\begin{aligned} \mathsf{Cov}[Y_{ki}, Y_{kj}] &= \mathsf{Cov}_{\boldsymbol{\gamma}}[\mathsf{E}_Y[Y_{ki}|\,\boldsymbol{\gamma}], \mathsf{E}_Y[Y_{kj}|\,\boldsymbol{\gamma}_k]] \\ & \qquad + \ \mathsf{E}_{\boldsymbol{\gamma}}[\mathsf{Cov}_Y[Y_{ki}, Y_{kj}|\,\boldsymbol{\gamma}_k]] \\ &= \Sigma_{\gamma,00} \ + \ \Sigma_{\gamma,01}(T_{ki} + T_{kj}) \ + \ \Sigma_{\gamma,11} T_{ki} T_{kj} \end{aligned}$$

- In contrast to the exchangeable structure induced by the random intercepts model, the random intercepts/slopes model permits:
  - ⋆ heteroskedasticity across the study units within a cluster
  - ⋆ covariance (correlation) to depend on the two study units under consideration

- Note both of these ‘extensions’ are specific to the covariate whose slope is permitted to vary across the clusters

- In principle, one could include any element of $\boldsymbol{X}_k$ in $\boldsymbol{Z}_k$ depending on
  - ⋆ what we believe about the structure of dependence
  - ⋆ the scientific interest(s)

- Visualization of K=10 clusters with n_k=5:



Special case #3: Random intercepts/serial dependence model

- The two special cases we've seen so far build flexibility in the first component of:

$$\begin{aligned} \text{Cov}[\boldsymbol{Y}_k | \ \boldsymbol{X}_k] &= \boldsymbol{\Sigma}_k(\boldsymbol{\alpha}) \\ &= \underbrace{\boldsymbol{Z}_k \boldsymbol{G}(\boldsymbol{\alpha}) \boldsymbol{Z}_k^T}_{\text{Random effects}} \quad + \quad \underbrace{\boldsymbol{R}_k(\boldsymbol{\alpha})}_{\text{Residual error}} \end{aligned}$$

  - ⋆ i.e. solely via the random effects

- An alternative strategy to introducing flexibility is to explicitly model the correlation structure in the $\boldsymbol{R}(\boldsymbol{\alpha})$
  - ⋆ i.e. move beyond

$$\boldsymbol{R}_k(\boldsymbol{\alpha}) \ = \ \sigma_\epsilon^2 I_{n_k}$$

    and permit the ϵ_{ki} to be correlated

- In a longitudinal study, one could do this by postulating that the residual error terms arise, in part, due to some (latent) stochastic process
  - ⋆ introduces *serial dependence*

- For example, the random intercepts/serial dependence model:

$$Y_{ki} = \boldsymbol{X}_{ki}^T \boldsymbol{\beta} + \gamma_{0k} + W_k(T_{ki}) + \epsilon_{ki}^*$$

  - ⋆ γ_{0k} is a cluster-specific random intercept
  - ⋆ assume $\mathsf{E}[\gamma_{0k}] = 0$ and $\mathsf{V}[\gamma_{0k}] = \sigma_\gamma^2$
  - ⋆ $W_k(T_{ki})$ is a serial dependence term
  - ⋆ assume $\mathsf{E}[\epsilon_{ki}^*] = 0$ and $\mathsf{V}[\epsilon_{ki}^*] = \sigma_{\epsilon^*}^2$

- Assume the stochastic process $W_k(\cdot)$ is mean zero and is characterized by its covariance function:

$$\mathsf{Cov}[W_k(T_{ki}),\ W_k(T_{kj})] = \sigma_W^2 \rho(U_{k,ij})$$

where $U_{k,ij} = |T_{ki} - T_{kj}|$

- Note, the ϵ^*_{ki} terms in this model don't have the same interpretation as in the two previous models
  - ⋆ capture variation in the error terms beyond that explained by $W_k(\cdot)$

- Under this model, assuming independence between γ_{0k} and $W_k(\cdot)$, for two specific study units

$$\begin{aligned} Y_{ki} &= \beta_0 + \beta_1 T_{ki} + \gamma_{0k} + W_k(T_{ki}) + \epsilon^*_{ki} \\ Y_{kj} &= \beta_0 + \beta_1 T_{kj} + \gamma_{0k} + W_k(T_{kj}) + \epsilon^*_{kj} \end{aligned}$$

we have:

$$\begin{aligned} \mathsf{V}[Y_{ki}] &= \mathsf{V}_\gamma[\mathsf{E}_Y[Y_{ki}|\ \gamma_{0k}]] + \mathsf{E}_\gamma[\mathsf{V}_Y[Y_{ki}|\ \gamma_{0k}]] \\ &= \sigma^2_\gamma + \sigma^2_W + \sigma^2_{\epsilon^*} \end{aligned}$$

$$\begin{aligned} \mathsf{Cov}[Y_{ki}, Y_{kj}] &= \mathsf{Cov}_\gamma[\mathsf{E}_Y[Y_{ki}|\ \gamma_{0k}], \mathsf{E}_Y[Y_{kj}|\ \gamma_{0k}]] \\ &\qquad + \mathsf{E}_\gamma[\mathsf{Cov}_Y[Y_{ki}, Y_{kj}|\ \gamma_{0k}]] \\ &= \sigma^2_\gamma + \sigma^2_W \rho(U_{k,ij}) \end{aligned}$$

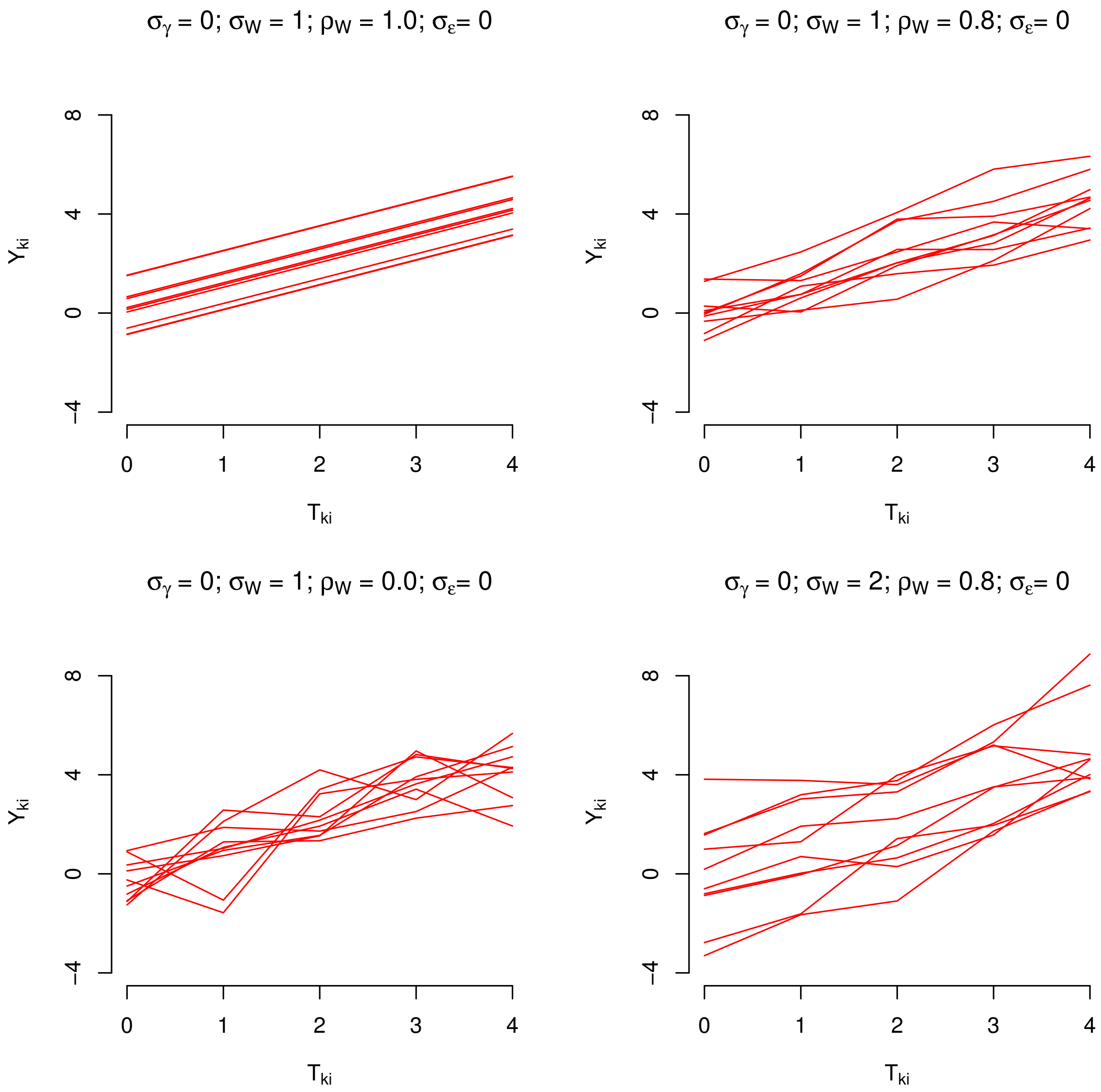
- Notice that the correlation between two study units is:

$$\text{Cor}[Y_{ki}, Y_{kj}] \;=\; \frac{\sigma_\gamma^2 \;+\; \sigma_W^2 \rho(U_{k,ij})}{\sigma_\gamma^2 \;+\; \sigma_W^2 \;+\; \sigma_{\epsilon^*}^2}$$

- Specific examples for the correlation function include:

$$\begin{array}{lll} \rho(U_{k,ij}) \;=\; \rho^{U_{kij}} & & \text{auto-regressive model} \\ \rho(U_{k,ij}) \;=\; \exp\{-U_{kij}/\text{range}\} & & \text{exponential spatial correlation} \end{array}$$

- Visualization of $K=10$ clusters with $n_k=5$ based on an auto-regressive correlation function for $W_k(\cdot)$:



Comment

- The random intercepts/slopes and the random intercepts/serial dependence models both structure $\Sigma_k(\boldsymbol{\alpha})$ such that correlation among study units depends, in part, on the timing of their measurement

- One way of thinking how to distinguish between the two models is that the:
  - ⋆ random intercepts/slopes model structures dependence *globally*
  - ⋆ random intercepts/serial dependence structures dependence *locally*

Estimation/inference

- Model components:
  - ⋆ response of the k^{th} cluster, given the fixed and random effects:

$$\boldsymbol{Y}_k = \boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{Z}_k\boldsymbol{\gamma}_k + \boldsymbol{\epsilon}_k$$

  - ⋆ random effects and residual error:

$$\begin{aligned} \mathsf{E}[\boldsymbol{\gamma}_k] &= \mathbf{0} & \mathsf{Cov}[\boldsymbol{\gamma}_k] &= \boldsymbol{G}(\boldsymbol{\alpha}) \\ \mathsf{E}[\boldsymbol{\epsilon}_k] &= \mathbf{0} & \mathsf{Cov}[\boldsymbol{\epsilon}_k] &= \boldsymbol{R}_k(\boldsymbol{\alpha}) \\ \mathsf{Cov}[\boldsymbol{\gamma}_k, \boldsymbol{\epsilon}_k] &= \mathbf{0} \end{aligned}$$

  - ⋆ responses across clusters are independent of eachother
- If we are to proceed on the basis of likelihood-based estimation/inference we need to specify a complete probability distribution for the data

The marginalized likelihood

- The most common way forward is to assume that and that

$$\begin{aligned}\boldsymbol{\gamma}_k &\sim \mathsf{MVN}_q(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\alpha}))\\ \boldsymbol{\epsilon}_k &\sim \mathsf{MVN}_{n_k}(\mathbf{0}, \boldsymbol{R}_k(\boldsymbol{\alpha}))\\ \boldsymbol{\gamma}_k &\perp\!\!\!\perp \boldsymbol{\epsilon}_k\end{aligned}$$

- Given these assumptions, we can write down the marginal distribution of the response for the k^{th} cluster as:

$$\begin{aligned}f_{\mathbf{Y}}(\boldsymbol{Y}_k|\ \boldsymbol{\beta}, \boldsymbol{\alpha}) &= \int f(\boldsymbol{Y}_k, \boldsymbol{\gamma}_k | \boldsymbol{\beta}, \boldsymbol{\alpha}) \partial \boldsymbol{\gamma}_k\\ &= \int f_{\mathbf{Y}|\boldsymbol{\gamma}}(\boldsymbol{Y}_k|\ \boldsymbol{\gamma}_k, \boldsymbol{\beta}, \boldsymbol{\alpha}) f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_k|\ \boldsymbol{\alpha}) \partial \boldsymbol{\gamma}_k\end{aligned}$$

  - ⋆ marginal with respect to the random effects
  - ⋆ a convolution of two multivariate Normal distributions

- Since the convolution of two Normal distributions is itself a Normal distribution, and using results we derived on slide 167, we can write that:

$$\boldsymbol{Y}_k |\, \boldsymbol{\beta}, \boldsymbol{\alpha} \ \sim\ \mathsf{MVN}_{n_k}(\boldsymbol{X}_k\boldsymbol{\beta},\ \boldsymbol{\Sigma}_k(\boldsymbol{\alpha}))$$

  where $\boldsymbol{\Sigma}_k(\boldsymbol{\alpha}) \ = \ \boldsymbol{Z}_k\boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_k^T + \boldsymbol{R}_k(\boldsymbol{\alpha})$

- Use this as a basis for an *intergated* or *marginal* likelihood that can be used to perform estimation/inference with respect to $(\boldsymbol{\beta}, \boldsymbol{\alpha})$:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \ = \ \prod_{k=1}^{K} f_{\boldsymbol{Y}}(\boldsymbol{Y}_k |\, \boldsymbol{\beta}, \boldsymbol{\alpha})$$

  - ⋆ assume that contributions from different clusters are independent

- From this, the log-likelihood is proportional to:

$$\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) \ \propto\ \sum_{k=1}^{K} \log|\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})| \ + \ \sum_{k=1}^{K} (\boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{\beta})^T \boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1} (\boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{\beta})$$

ML estimation of $\boldsymbol{\beta}$

- Differentiating with respect to $\boldsymbol{\beta}$ yields the score:

$$U_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\alpha}) \;=\; \frac{\partial}{\partial\boldsymbol{\beta}}\ell(\boldsymbol{\beta},\boldsymbol{\alpha}) \;=\; \sum_{k=1}^{K}\boldsymbol{X}_k^T\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1}(\boldsymbol{Y}_k-\boldsymbol{X}_k\boldsymbol{\beta})$$

  which yields the familiar MLE for $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \;=\; \left(\sum_{k=1}^{K}\boldsymbol{X}_k^T\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1}\boldsymbol{X}_k\right)^{-1}\left(\sum_{k=1}^{K}\boldsymbol{X}_k^T\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1}\boldsymbol{Y}_k\right)$$

- The covariance of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is based on the inverse of the Fisher information matrix:

$$\mathcal{I}_{\boldsymbol{\beta}}(\boldsymbol{\beta},\boldsymbol{\alpha}) \;=\; -\mathsf{E}\left[\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\ell(\boldsymbol{\beta},\boldsymbol{\alpha})\right] \;=\; \sum_{k=1}^{K}\boldsymbol{X}_k^T\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1}\boldsymbol{X}_k$$

  ⋆ denoted this by $\boldsymbol{A}(\boldsymbol{\alpha})$ when discussing least squares estimation

- Notice that the MLE is equivalent to the WLS estimator with $\boldsymbol{W}$ taken to be $\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1}$
  - ⋆ we could therefore use the theory of WLS to perform estimation/ inference for $\boldsymbol{\beta}$
  - ⋆ since $\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is a linear combination of the $\boldsymbol{Y}_k$, we have:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \sim \text{MVN}_p(\boldsymbol{\beta},\ \boldsymbol{A}(\boldsymbol{\alpha})^{-1})$$

  - ⋆ since $\boldsymbol{\alpha}$ is unknown, we could plug-in any consistent estimator and appeal to Slutsky's Theorem to see that:

$$\boldsymbol{A}(\widehat{\boldsymbol{\alpha}})^{-1/2}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\alpha}}) - \boldsymbol{\beta}) \longrightarrow \text{MVN}_p(\mathbf{0},\ \boldsymbol{I})$$

    as $K \longrightarrow \infty$
  - ⋆ note, it is not sufficient for $n_k \longrightarrow \infty$ for fixed K

ML estimation of $\boldsymbol{\alpha}$

- Differentiating the log-likelihood with respect to $\boldsymbol{\alpha}$ yields:

$$
\begin{aligned}
U_{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \frac{\partial}{\partial \boldsymbol{\alpha}} \ell(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\
&= \sum_{k=1}^{K} \left[(\boldsymbol{Y}_k - \boldsymbol{X}_k \boldsymbol{\beta})^T \boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1} \frac{\partial \boldsymbol{\Sigma}_k}{\partial \boldsymbol{\alpha}} \boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1} (\boldsymbol{Y}_k - \boldsymbol{X}_k \boldsymbol{\beta}) \right. \\
&\qquad \left. - \, \mathsf{trace} \left(\boldsymbol{\Sigma}_k^{-1} \frac{\partial \boldsymbol{\Sigma}_k}{\partial \boldsymbol{\alpha}} \right) \right]
\end{aligned}
$$

- In general one won't be able to solve $U_{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{0}$ analytically

- However one proceeds, the solution $\widehat{\boldsymbol{\alpha}}$ must satisfy the constraint that the resulting covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\alpha}})$ is positive definite
  - ⋆ several computation methods for parameterizing $\boldsymbol{\Sigma}(\boldsymbol{\alpha})$ so that this constraint is satisfied
  - ⋆ e.g. Cholesky decomposition

REML estimation

- Recall that, in general, the MLE for variance components in a multivariate Normal exhibit small sample bias
  - ⋆ intuition is that the standard ML estimator doesn't acknowledge the fact that the mean (i.e. $\boldsymbol{\beta}$) is estimated
  - ⋆ implications for inference in small samples

- When considering estimation/inference for general linear models for dependent data we derived the REML estimator as an alternative

- REML can also be used here

- Operationally, the strategy involves the use of the linear transformation:

$$\boldsymbol{Y} \Rightarrow (\boldsymbol{Z}, \widehat{\boldsymbol{\beta}})$$

  where $\boldsymbol{Z} = B^T\boldsymbol{Y}$ with B the $N \times (N-p)$ matrix defined by the requirements that $BB^T = A$ and $B^TB = \boldsymbol{I}$.

- Exploiting the facts that E$[\boldsymbol{Z}] = \mathbf{0}$ and Cov$[\boldsymbol{Z}, \widehat{\boldsymbol{\beta}}] = \mathbf{0}$, the pdf of $\boldsymbol{Z}$, expressed as a function of $\boldsymbol{Y}$, is proportional to:

$$\begin{aligned}\frac{f_{\boldsymbol{Y}}(\boldsymbol{Y})}{f(\widehat{\boldsymbol{\beta}})} &= (2\pi)^{-(N-p)/2}|\boldsymbol{\Sigma}(\boldsymbol{\alpha})|^{-1/2}|\boldsymbol{X}^T\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}\boldsymbol{X}|^{-1/2} \\ &\quad \times \ \exp\left\{-\frac{1}{2}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\right\}\end{aligned}$$

- The REML estimator of $\boldsymbol{\Sigma}$, therefore, maximizes the so-called *restricted log-likelihood*:

$$\begin{aligned}\ell^*(\boldsymbol{\alpha}) \ \propto\ &- \ \log|\boldsymbol{\Sigma}| \ - \ \log|\boldsymbol{X}^T\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}\boldsymbol{X}| \\ &- \ (\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})^T\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}(\boldsymbol{Y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\end{aligned}$$

Inference

- From an inferential perspective, one might be interested in answering questions regarding the mean model and/or the dependence model

- Mean model:
  - ⋆ formally evaluate the association between some exposure of interest and the response
  - ⋆ compare competing specifications of some association
  - ⋆ investigate potential effect modification

- Dependence model:
  - ⋆ explain the nature of random variation in the data
  - ⋆ valid inference for the fixed effects requires a ‘correct’ structure for $\mathbf{\Sigma}_k(\boldsymbol{\alpha})$
  - ⋆ avoid over-parameterization which can lead to inefficient inference for the fixed effects

Inference for the mean model

- Consider testing the fixed effects in nested linear mixed models:

$$H_0 : \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix} \quad \text{versus} \quad H_1 : \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$$

- Usual Wald test will be valid, regardless of whether ML or REML was used
- If ML was used, one could also use a likelihood ratio test
- If REML was used, a likelihood ratio test will not be valid
  - ⋆ $\boldsymbol{X}_k\boldsymbol{\beta}$ under H_0 is not the same as $\boldsymbol{X}_k\boldsymbol{\beta}$ under H_1
  - ⋆ correspondingly different $\boldsymbol{Z} = B^T\boldsymbol{Y}$ under H_0 and H_1
  - ⋆ REML likelihoods are based on different sets of observations
  - ⋆ differences in REML deviances are not meaningful

Inference for the dependence structure

- Consider testing whether the random slopes in a random intercepts/slopes model are ‘necessary’
  - ⋆ i.e. compare the fit based on random intercepts/slopes model:

$$Y_{ki} \;=\; \boldsymbol{X}_{ki}^T\boldsymbol{\beta} \;+\; \gamma_{0k} \;+\; T_{ki}\gamma_{1k} \;+\; \epsilon_{ki},$$

    to that based on the random intercepts model:

$$Y_{ki} \;=\; \boldsymbol{X}_{ki}^T\boldsymbol{\beta} \;+\; \gamma_{0k} \;+\; \epsilon_{ki}$$

- Recall the covariance structure for the random intercepts and slopes:

$$\text{Cov}[\boldsymbol{\gamma}_k] \;=\; \boldsymbol{G}(\boldsymbol{\alpha}) \;=\; \begin{bmatrix} \Sigma_{\gamma,00} & \Sigma_{\gamma,01} \\ \Sigma_{\gamma,01} & \Sigma_{\gamma,11} \end{bmatrix}$$

- We can therefore structure the hypotheses for this question as:

$$H_0 : \boldsymbol{G}(\boldsymbol{\alpha}) = \begin{bmatrix} \Sigma_{\gamma,00} & 0 \\ 0 & 0 \end{bmatrix}$$

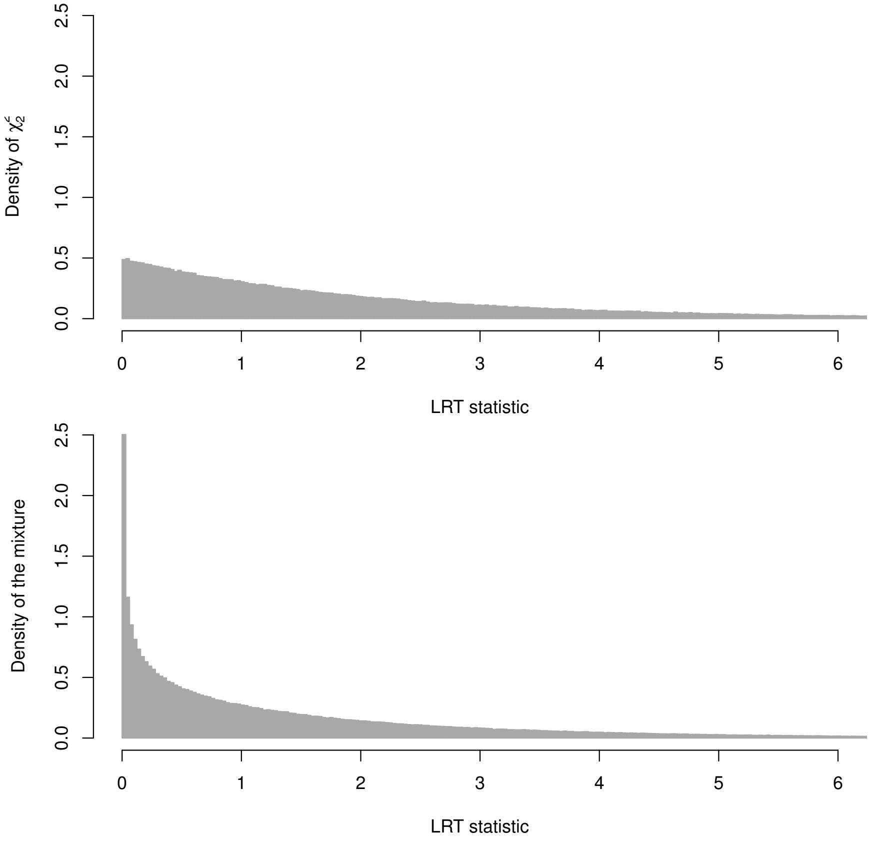
versus

$$H_1 : \boldsymbol{G}(\boldsymbol{\alpha}) = \begin{bmatrix} \Sigma_{\gamma,00} & \Sigma_{\gamma,01} \\ \Sigma_{\gamma,01} & \Sigma_{\gamma,11} \end{bmatrix}$$

- Notice that under the null, $\Sigma_{\gamma,11} = 0$
  - ⋆ $\boldsymbol{G}(\boldsymbol{\alpha})$ under H_0 is the boundary of the parameter space defined by the alternative
  - ⋆ violation of a standard assumption used to establish the (asymptotic) χ^2 distribution of the likelihood ratio test statistic

- In general, when testing variance components, the asymptotic sampling distribution of the LRT statistic is a *mixture of* χ^2 *distributions*
    - ⋆ nature of the mixture depends on the hypotheses one is testing

- For H_0 and H_1 above, the correct sampling distribution is a 50:50 mixture of a χ_2^2 distribution and a χ_1^2 distribution



- See that the naïve use of a χ^2_2 distribution will lead to overly conservative inference
  - ⋆ for a given value of the test statistic, the p-value based on a χ^2_2 distribution will be <u>bigger</u> than the p-value based on the correct mixture
  - ⋆ fail to reject H_0 when one should reject it
  - ⋆ leads to an incorrect over-simplification of the model for the dependence structure

- Additional details can be found in:
  - ⋆ Self and Liang (JASA, 1987)
  - ⋆ Stram and Lee (Biometrics, 1994)

- Finally, note that, in contrast to likelihood-based inference for $\boldsymbol{\beta}$, one can perform valid likelihood-based inference for $\boldsymbol{\alpha}$ whether ML or REML is used
  - ⋆ $\boldsymbol{X}_k\boldsymbol{\beta}$ is the same under H_0 and H_1
  - ⋆ REML likelihoods are based on the same set of observations

Empirical Bayes

- So far estimation/inference for ($\boldsymbol{\beta}$, $\boldsymbol{\alpha}$) has been based on a marginalized likelihood in which the random effects have been integrated out

- In some settings, one might be interested in the random effects themselves
  - ⋆ visualization of heterogeneity across clusters
  - ⋆ characterization of cluster-specific associations
  - ⋆ profiling and ranking

- If ($\boldsymbol{\beta}$, $\boldsymbol{\alpha}$) were known, one could proceed within the Bayesian paradigm for the 'unknown' $\boldsymbol{\gamma}_k$:
  - ⋆ take the conditional distribution of the data as the 'likelihood'
  - ⋆ take the distributional assumptions for the random effects as the 'prior'
  - ⋆ the posterior of $\boldsymbol{\gamma}_k$ given the data $\boldsymbol{Y}_k$:

$$\pi(\boldsymbol{\gamma}_k | \, \boldsymbol{Y}_k, \boldsymbol{\beta}, \boldsymbol{\alpha}) \; \propto \; \mathcal{L}(\boldsymbol{\gamma}_k; \boldsymbol{Y}_k, \boldsymbol{\beta}, \boldsymbol{\alpha}) \; \times \; \pi(\boldsymbol{\gamma}_k | \, \boldsymbol{\alpha})$$

- As an ‘estimate’ of $\boldsymbol{\gamma}_k$, one might compute and report the mean of the posterior distribution:

$$\widetilde{\boldsymbol{\gamma}}_k = \mathsf{E}[\boldsymbol{\gamma}_k | \, \boldsymbol{Y}_k, \boldsymbol{\beta}, \boldsymbol{\alpha}]$$

- For a given form of this expectation, one could plug in estimates of ($\boldsymbol{\beta}$, $\boldsymbol{\alpha}$) based on ML or REML

- Referred to as the *empirical Bayes* estimator
  - ⋆ empirical in the sense that the data was used to inform the ‘prior’
  - ⋆ i.e. we plugged $\widehat{\boldsymbol{\alpha}}$ into $\pi(\boldsymbol{\gamma}_k | \, \boldsymbol{\alpha})$

Empirical Bayes in a simple setting

- Consider the simple setting in which there are no covariates:

$$Y_{ki} \;=\; \beta_{0k} \;+\; \epsilon_{ki}$$

  - ⋆ $\beta_{0k} \sim \text{Normal}(\beta_0,\, \tau^2)$
  - ⋆ $\epsilon_{ki} \sim \text{Normal}(0,\, \sigma^2)$

- The induced marginal distribution of the response vector is:

$$\begin{bmatrix} Y_{k1} \\ Y_{k2} \\ \vdots \\ Y_{kn_k} \end{bmatrix} \;\sim\; \text{MVN}_{n_k}\left(\begin{bmatrix} \beta_0 \\ \beta_0 \\ \vdots \\ \beta_0 \end{bmatrix}, \begin{bmatrix} \tau^2+\sigma^2 & \tau^2 & \ldots & \tau^2 \\ \tau^2 & \tau^2+\sigma^2 & \ldots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \ldots & \tau^2+\sigma^2 \end{bmatrix} \right)$$

- It is relatively straightforward to show that the ML estimator of β_0, the population mean, is simply the WLS estimator:

$$\hat{\beta}_0 \;=\; \frac{\sum_k w_k \overline{Y}_k}{\sum_k w_k}$$

  where $w_k = \tau^2 + \sigma^2/n_k$

  ⋆ obtain estimates of σ^2 and τ^2 via ML or REML

- Now consider the joint distribution of the cluster-specific mean, $\overline{Y}_k$ and β_{0k}:

$$\begin{bmatrix} \overline{Y}_k \\ \beta_{0k} \end{bmatrix} \sim \text{MVN}_2\left(\begin{bmatrix} \beta_0 \\ \beta_0 \end{bmatrix}, \begin{bmatrix} \tau^2 + \sigma^2/n_k & \tau^2 \\ \tau^2 & \tau^2 \end{bmatrix} \right)$$

- The corresponding conditional distribution of $\beta_{0k} | \, \overline{Y}_k$ is a Normal with mean

$$\begin{aligned} \mathsf{E}[\beta_{0k} | \, \overline{Y}_k, \beta_0, \sigma^2, \tau^2] &= \beta_0 \; + \; \frac{\tau^2}{\tau^2 + \sigma^2/n_k} (\overline{Y}_k - \beta_0) \\ &= \frac{\sigma^2/n_k}{\tau^2 + \sigma^2/n_k} \beta_0 \; + \; \frac{\tau^2}{\tau^2 + \sigma^2/n_k} \overline{Y}_k \end{aligned}$$

  - ⋆ weighted average of the population mean and the cluster-specific mean of the responses
  - ⋆ as n_k gets large, more weight is attached to the sample mean
  - ⋆ also depends on the relative degree of between- versus within-cluster variation

- Operationally one would plug-in the estimates of β_0, σ^2 and τ^2, we get the empirical Bayes estimate of β_{0k}:

$$\tilde{\beta}_{0k} \;=\; \frac{\hat{\sigma}^2/n_k}{\hat{\tau}^2 + \hat{\sigma}^2/n_k} \hat{\beta}_0 \; + \; \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2/n_k} \overline{Y}_k$$

Empirical Bayes for linear mixed effects models

- Now consider the linear mixed effects model:

$$\begin{aligned}\boldsymbol{Y}_k &= \boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{Z}_k\boldsymbol{\gamma}_k + \boldsymbol{\epsilon}_k \\ \boldsymbol{\gamma}_k &\sim \text{MVN}_q(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\alpha})) \\ \boldsymbol{\epsilon}_k &\sim \text{MVN}_{n_k}(\mathbf{0}, \boldsymbol{R}_k(\boldsymbol{\alpha})) \\ \boldsymbol{\gamma}_k &\perp\!\!\!\perp \boldsymbol{\epsilon}_k\end{aligned}$$

  for which the induced marginal distribution of $\boldsymbol{Y}_k$ is a multivariate Normal:

$$\boldsymbol{Y}_k \sim \text{MVN}_{n_k}(\boldsymbol{X}_k\boldsymbol{\beta},\ \boldsymbol{\Sigma}_k(\boldsymbol{\alpha}))$$

  where $\boldsymbol{\Sigma}_k(\boldsymbol{\alpha}) = \boldsymbol{Z}_k\boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_k^T + \boldsymbol{R}_k(\boldsymbol{\alpha})$

- Suppose we obtain estimates of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ via ML or REML

- Now consider the (cluster-specific) joint distribution of $\boldsymbol{Y}_k$ and $\boldsymbol{\gamma}_k$:

$$\begin{bmatrix} \boldsymbol{Y}_k \\ \boldsymbol{\gamma}_k \end{bmatrix} \sim \text{MVN}_{n_k+q}\left(\begin{bmatrix} \boldsymbol{X}_k\boldsymbol{\beta} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_k(\boldsymbol{\alpha}) & \boldsymbol{Z}_k\boldsymbol{G}(\boldsymbol{\alpha}) \\ \boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_k^T & \boldsymbol{G}(\boldsymbol{\alpha}) \end{bmatrix} \right)$$

since

$$\begin{aligned} \text{Cov}[\boldsymbol{Y}_k,\ \boldsymbol{\gamma}_k] &= \text{Cov}[\boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{Z}_k\boldsymbol{\gamma}_k + \boldsymbol{\epsilon}_k,\ \boldsymbol{\gamma}_k] \\ &= \text{Cov}[\boldsymbol{X}_k\boldsymbol{\beta},\ \boldsymbol{\gamma}_k] + \text{Cov}[\boldsymbol{Z}_k\boldsymbol{\gamma}_k,\ \boldsymbol{\gamma}_k] + \text{Cov}[\boldsymbol{\epsilon}_k,\ \boldsymbol{\gamma}_k] \\ &= \boldsymbol{0} + \boldsymbol{Z}_k\text{Cov}[\boldsymbol{\gamma}_k,\ \boldsymbol{\gamma}_k] + \boldsymbol{0} \\ &= \boldsymbol{Z}_k\boldsymbol{G}(\boldsymbol{\alpha}) \end{aligned}$$

- The induced conditional distribution of $\boldsymbol{\gamma}_k |\ \boldsymbol{Y}_k$ is:

$$\boldsymbol{\gamma}_k |\ \boldsymbol{Y}_k \sim \text{MVN}_q(\boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_k^T\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1}(\boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{\beta}),\ \boldsymbol{\Sigma}_k^*(\boldsymbol{\alpha}))$$

where $\boldsymbol{\Sigma}_k^*(\boldsymbol{\alpha}) = \boldsymbol{G}(\boldsymbol{\alpha}) - \boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_k^T\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})^{-1}\boldsymbol{Z}_k\boldsymbol{G}(\boldsymbol{\alpha})$

- Consequently, the empirical Bayes estimator of $\boldsymbol{\gamma}_k$ is:

$$\widetilde{\boldsymbol{\gamma}}_k = \boldsymbol{G}(\widehat{\boldsymbol{\alpha}})\boldsymbol{Z}_k^T\boldsymbol{\Sigma}_k(\widehat{\boldsymbol{\alpha}})^{-1}(\boldsymbol{Y}_k - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}})$$

  ⋆ simply plug in the estimates (ML or REML) of ($\boldsymbol{\beta}$, $\boldsymbol{\alpha}$)

- Can also examine the fitted values of the response based on the empirical Bayes estimator:

$$\begin{aligned}\widehat{\boldsymbol{Y}}_k &= \boldsymbol{X}_k\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_k\widetilde{\boldsymbol{\gamma}}_k\\ &= \boldsymbol{X}_k\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_k\left[\boldsymbol{G}(\widehat{\boldsymbol{\alpha}})\boldsymbol{Z}_k^T\boldsymbol{\Sigma}_k(\widehat{\boldsymbol{\alpha}})^{-1}(\boldsymbol{Y}_k - \boldsymbol{X}_k\widehat{\boldsymbol{\beta}})\right]\\ &= [\boldsymbol{I}_{n_k} - \boldsymbol{Z}_k\boldsymbol{G}(\widehat{\boldsymbol{\alpha}})\boldsymbol{Z}_k^T\boldsymbol{\Sigma}_k(\widehat{\boldsymbol{\alpha}})^{-1}]\boldsymbol{X}_k\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_k\boldsymbol{G}(\widehat{\boldsymbol{\alpha}})\boldsymbol{Z}_k^T\boldsymbol{\Sigma}_k(\widehat{\boldsymbol{\alpha}})^{-1}\boldsymbol{Y}_k\end{aligned}$$

  ⋆ weighted average of the estimated population profile, $\boldsymbol{X}_k\widehat{\boldsymbol{\beta}}$, and the observed data, $\boldsymbol{Y}_k$.

Fitting models in `R`

- There are many ways of fitting linear mixed effects models in `R`

- Two functions/packages written by Doug Bates and colleagues are:
  - ⋆ `lme()` function in the `nlme` package
  - ⋆ `lmer()` function in the `lme4` package

- Unfortunately, no single function/package has the functionality to fit all forms of a linear mixed effects model that one could be interested in
  - ⋆ e.g. `lme()` permits direct specification of structure for the off-diagonal elements of $\boldsymbol{R}_k(\boldsymbol{\alpha})$, whereas `lmer()` does not (at least as far as I can tell)
  - ⋆ e.g. `lmer()` permits non-nested levels of clustering, whereas `lme()` does not (at least as far as I can tell)

- The basic call to `lme()` is has the following elements:

`fixed`	Specification for the regression model, $\boldsymbol{X}_k\boldsymbol{\beta}$
`random`	Specification for the random effects model, $\boldsymbol{Z}_k\boldsymbol{\gamma}_k$
`correlation`	Specification for serial dependence in $\boldsymbol{R}_k(\boldsymbol{\alpha})$
`weights`	Specification for heteroskedasticity in $\boldsymbol{R}_k(\boldsymbol{\alpha})$
`method`	Indicator of whether ML or REML is used

```
> ##
> ?lme                  ## Help on specifying a linear mixed model
> ?corClasses           ## Help on specifying serial dependence
> ?varClasses           ## Help on specifying heteroskedasticity
```

- Additional resources:
  - ⋆ Pinheiro J. and Bates D. *Mixed-Effects Models in S and S-PLUS* (2006)
  - ⋆ Galecki A. and Tomasz B. *Linear mixed-effects models using R: A step-by-step approach* (2013).

Dental growth data

- Fit a series of models with the same mean structure:

$$\mathsf{E}[Y_{ki}] \;=\; \beta_0 \;+\; \beta_1 A^*_{ki} \;+\; \beta_1 G_k \;+\; \beta_3 A^*_{ki} G_k$$

  - ⋆ where $A^*_{ki} = A_{ki} - 8$

- Consider a range of specifications for dependence:

$$\boldsymbol{\Sigma}_k(\boldsymbol{\alpha}) \;=\; \boldsymbol{Z}_k \boldsymbol{G}(\boldsymbol{\alpha}) \boldsymbol{Z}_k^T + \boldsymbol{R}_k(\boldsymbol{\alpha})$$

```
> ##
> load("Growth.RData")
> growth$ageStar <- growth$age - 8
> library(nlme)
>
> ## Naive - independent errors
> ##
> fit0.ML <- glm(length ~ ageStar * gender, data=growth, family=gaussian)
```

```
> ## Random intercepts + independent errors
> ##
> fit1.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ 1 | id),
+                data=growth, method="ML")
>
> ## Independent random intercepts/slopes + independent errors
> ##  * two subject-specific random effects are independent
> ##
> fit2.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ ageStar | id, pdClass="pdDiag"),
+                data=growth, method="ML")
>
> ## Correlated random intercepts/slopes + independent errors
> ##  * two subject-specific random effects are correlated
> ##
> fit3.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ ageStar | id),
+                data=growth, method="ML")
```

```
> ## Random intercepts + auto-regressive errors
> ##
> fit4.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ 1 | id),
+                correlation=corAR1(form= ~ ageStar| id),
+                data=growth, method="ML")
>
> ## Random intercepts + exponential spatial errors
> ##
> fit5.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ 1 | id),
+                correlation=corExp(form= ~ ageStar| id),
+                data=growth, method="ML")
>
> ## Random intercepts + (exponential spatial with a ‘nugget’)
> ##  - see the help file for corExp()
> ##
> fit6.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ 1 | id),
+                correlation=corExp(form= ~ ageStar| id, nugget=TRUE),
+                data=growth, method="ML")
```

```
> ## Random intercepts + independent errors
> ## * heteroskedasticity across age
> ##
> fit7.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ 1 | id),
+                weights=varIdent(form= ~1 | ageStar),
+                data=growth, method="ML")
>
> ## Random intercepts + independent errors
> ## * heteroskedasticity across gender
> ##
> fit8.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ 1 | id),
+                weights=varIdent(form= ~1 | gender),
+                data=growth, method="ML")
```

- Output:

```
> ##
> summary(fit0.ML)
...
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         21.2091     0.5700  37.207  < 2e-16 ***
ageStar              0.4795     0.1523   3.148  0.00217 **
gendermale           1.4909     0.7505   1.987  0.04970 *
ageStar:gendermale   0.3205     0.2006   1.598  0.11326
...
(Dispersion parameter for gaussian family taken to be 5.105977)

    Null deviance: 905.84  on 103  degrees of freedom
Residual deviance: 510.60  on 100  degrees of freedom
AIC: 470.62
...
> sqrt(summary(fit0.ML)$dispersion)
[1] 2.259641
```

```
> ## Random intercepts + independent errors
> ##
> summary(fit1.ML)
Linear mixed-effects model fit by maximum likelihood
 Data: growth
       AIC      BIC    logLik
  426.1665 442.0329 -207.0833

Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    1.731043 1.383142

Fixed effects: length ~ ageStar * gender
                       Value Std.Error DF  t-value p-value
(Intercept)        21.209091 0.6402482 76 33.12636  0.0000
ageStar             0.479545 0.0950982 76  5.04264  0.0000
gendermale          1.490909 0.8429259 24  1.76873  0.0896
ageStar:gendermale  0.320455 0.1252026 76  2.55949  0.0125
...
Number of Observations: 104
Number of Groups: 26
```

```
> ## Independent random intercepts/slopes + independent errors
> ##
> summary(fit2.ML)
...
Random effects:
 Formula: ~ageStar | id
 Structure: Diagonal
        (Intercept)  ageStar Residual
StdDev:    1.692609 0.142729 1.332001
...
>
> ## Correlated random intercepts/slopes + independent errors
> ##
> summary(fit3.ML)
...
Random effects:
 Formula: ~ageStar | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 1.7236512 (Intr)
ageStar     0.1543159 -0.093
Residual    1.3245101
....
```

```
> ## Random intercepts + auto-regressive errors
> ##
> summary(fit4.ML)
...
Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    1.731043 1.383142

Correlation Structure: ARMA(1,0)
 Formula: ~ageStar | id
 Parameter estimate(s):
Phi1
   0
...
```

```
> ## Random intercepts + exponential spatial errors
> ##
> summary(fit5.ML)
...
Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    1.731043 1.383142

Correlation Structure: Exponential spatial correlation
 Formula: ~ageStar | id
 Parameter estimate(s):
    range
0.1079813
...
>
> ## Correlation between two observation 1 time unit apart
> ##
> range <- 0.1079813
> exp(-1/range)
[1] 9.507362e-05
```

```
> ## Random intercepts + (exponential spatial with a ‘nugget’)
> ##
> summary(fit6.ML)
...
Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    1.731043 1.383142

Correlation Structure: Exponential spatial correlation
 Formula: ~ageStar | id
 Parameter estimate(s):
     range     nugget
0.11013210 0.09606834
...
>
> ## Correlation between two observation 1 time unit apart
> ##
> range  <- 0.11013210
> nugget <- 0.09606834
> (1-nugget) * exp(-1/range)
[1] 0.0001029769
```

```
> ## Random intercepts + independent errors
> ## * heteroskedasticity across age
> summary(fit7.ML)
...
Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    1.722349 1.635319

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | ageStar
 Parameter estimates:
        0         2         4         6
1.0000000 0.7392948 0.9315600 0.6922482
...
```

```
> ## Random intercepts + independent errors
> ## * heteroskedasticity across gender
> summary(fit8.ML)
...
Random effects:
 Formula: ~1 | id
        (Intercept)  Residual
StdDev:    1.766799 0.7694708

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | gender
 Parameter estimates:
  female     male
1.000000 2.193266
...
```

- Comparison of model fit:

Dependence model	**log-Like**	**AIC**
0. Independence	-230.3115	470.6231
1. Random intercepts + inde. errors	-207.0833	426.1665
2. Independent random intercepts/slopes + inde. errors	-206.7674	427.5348
3. Correlated random intercepts/slopes + inde. errors	-206.7540	429.5081
4. Random intercepts + AR errors	-207.0833	428.1665
5. Random intercepts + ES errors	-207.0833	428.1665
6. Random intercepts + (ES with a 'nugget')	-207.0833	430.1665
7. Random intercepts + heteroske inde. errors (age)	-206.0505	430.1009
8. Random intercepts + heteroske inde. errors (gender)	-197.0882	408.1765

⋆ seems clear that model 8 provides the best fit (from among those presented, at least)

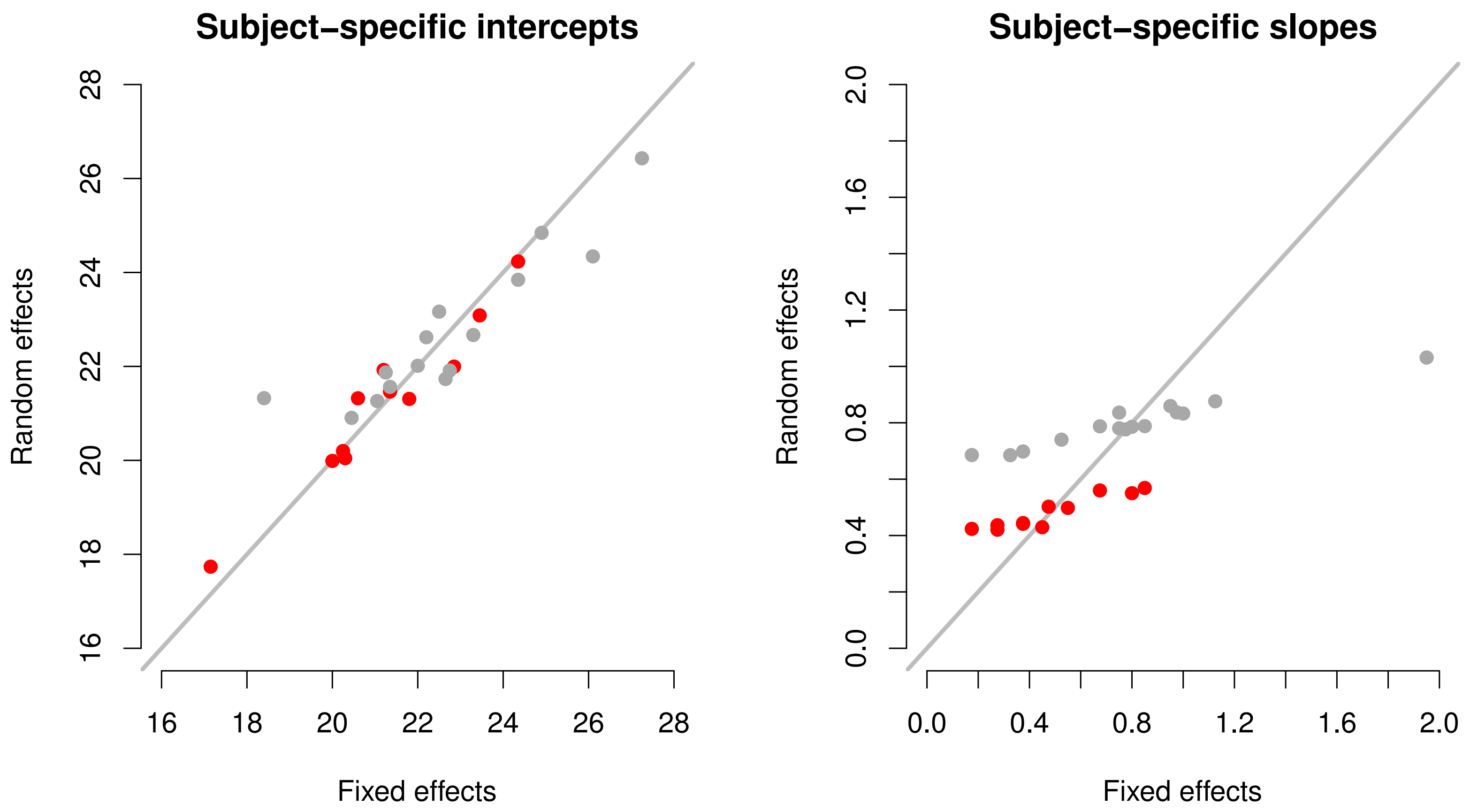
- Results for:

$$\mathsf{E}[Y_{ki}] \;=\; \beta_0 \;+\; \beta_1 A^*_{ki} \;+\; \beta_1 G_k \;+\; \beta_3 A^*_{ki} G_k$$

Dependence	**Estimate**				**Standard error**			
model	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
0.	21.21	0.48	1.49	0.32	0.570	0.152	0.750	0.201
1.	21.21	0.48	1.49	0.32	0.640	0.095	0.843	0.125
2.	21.21	0.48	1.49	0.32	0.623	0.102	0.820	0.134
3.	21.21	0.48	1.49	0.32	0.630	0.103	0.830	0.135
4.	21.21	0.48	1.49	0.32	0.640	0.095	0.843	0.125
5.	21.21	0.48	1.49	0.32	0.640	0.095	0.843	0.125
6.	21.21	0.48	1.49	0.32	0.640	0.095	0.843	0.125
7.	21.22	0.48	1.37	0.35	0.650	0.092	0.855	0.122
8.	21.21	0.48	1.49	0.32	0.578	0.053	0.830	0.113

⋆ standard error estimates are notably smaller under model 8

- Empirical Bayes estimates of the random effects estimates from a random intercept/slopes model
  - ⋆ compare estimates with those from $K = 26$ separate regressions



  - ⋆ see structure imposed by the model

- Finally consider the impact of using REML for model 8

```
>
> fit8.REML <- lme(fixed=length ~ ageStar * gender,
+                  random=reStruct(~ 1 | id),
+                  weights=varIdent(form= ~1 | gender),
+                  data=growth, method="REML")
>
> summary(fit8.REML)
Linear mixed-effects model fit by REML
...
Random effects:
 Formula: ~1 | id
        (Intercept)  Residual
StdDev:    1.844801 0.7813232

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | gender
 Parameter estimates:
  female     male
1.000000 2.185356
...
```

	ML		REML	
	Est	SE	Est	SE
Fixed effects				
Intercept, β_0	21.21	0.578	21.21	0.590
Main effect for age, β_1	0.48	0.053	0.48	0.053
Main effect for gender, β_2	1.49	0.830	1.49	0.843
Interaction, β_3	0.32	0.113	0.32	0.112
Variance components				
SD of random intercepts	1.767		1.845	
SD for errors				
females	0.769		0.781	
males	1.688		1.707	

⋆ differences don't appear to be substantial

MACS CD4+ cell count data

- Consider the mean model:

$$\begin{aligned} \text{E}[\text{CD4}_{ki}] &= \beta_0 + \beta_1\text{time}_{ki} + \beta_2\text{age}_{ki} + \beta_3\text{smoker}_{ki} \\ &\qquad + \beta_4\text{drug}_{ki} + \beta_5\text{cesd-CS}_{ki} + \beta_5\text{cesd-LO}_{ki} \end{aligned}$$

  based on the following data manipulations:

```
> ##
> load("MACS.RData")
> macs$smoker <- as.numeric(macs$packs > 0)
> macs$age    <- macs$age / 5
> macs$cesd   <- (macs$cesd - 10) / 5
```

  - ⋆ restricted to K=266 participants those who seroconverted with the 'pre' measurement no more than 6 months prior to seroconversion
  - ⋆ `cesd-CS` and `cesd-LO` are the cross-sectional and longitudinal forms of CESD we've considered previously

- Fit seven of the models for the dependence structure considered for the dental growth data
  - ⋆ for simplicity do not consider heteroskedasticity

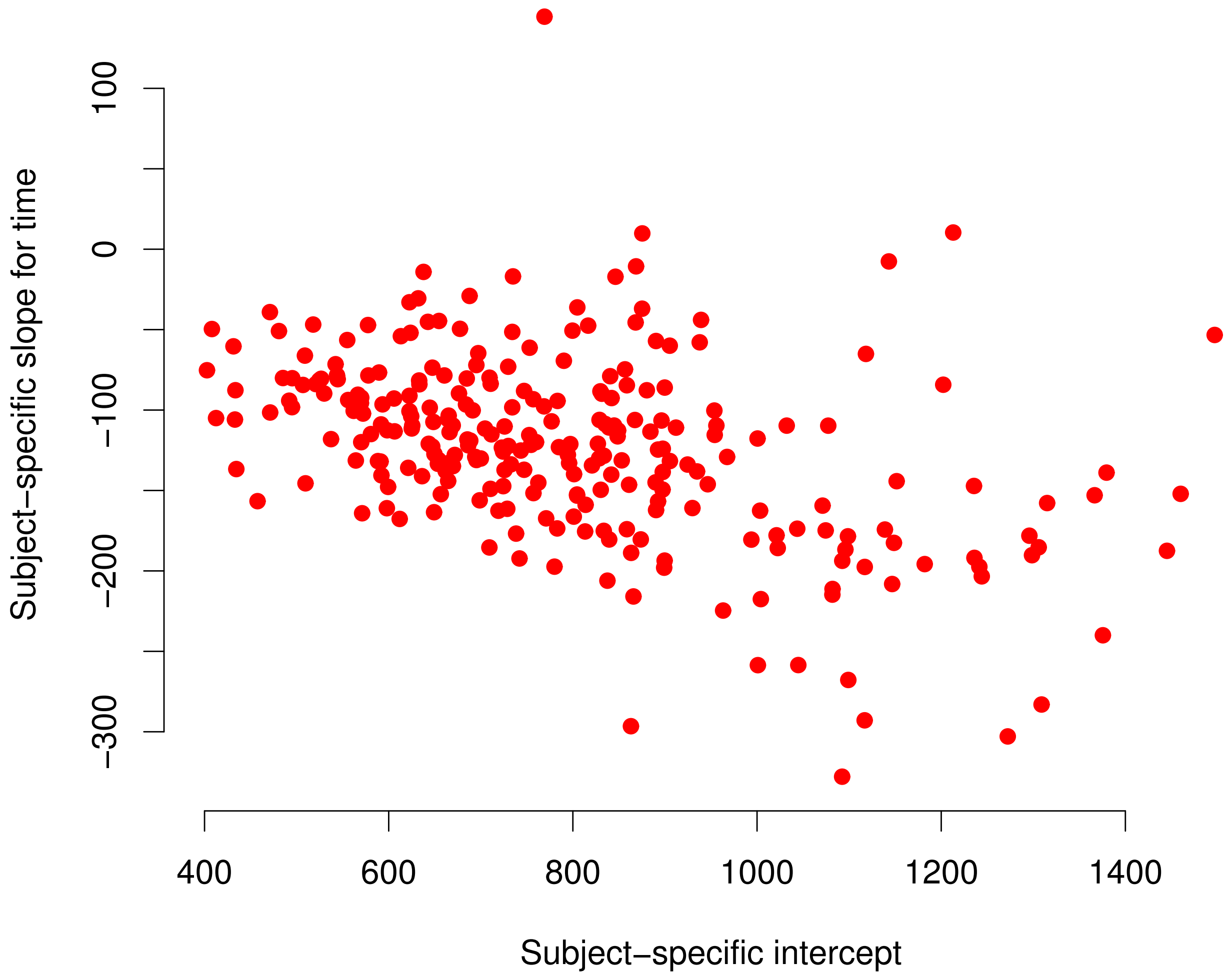
Dependence model	**log-Like**	**AIC**
0. Naïve independence	-10,956	21,927
1. Random intercepts + inde. errors	-10,705	21,428
2. Independent random intercepts/slopes + inde. errors	-10,674	21,368
3. Correlated random intercepts/slopes + inde. errors	-10,663	21,349
4. Random intercepts + AR errors	-10,832	21,684
5. Random intercepts + ES errors	-10,677	21,375
6. Random intercepts + (ES with a 'nugget')	-10,667	21,357

  - ⋆ indications that dependence models 3 and 6 provide the best fits to the data

```
> ## Correlated random intercepts/slopes + independent errors
> ##
> summary(fit3.ML)
...
Random effects:
 Formula: ~time | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 248.87779 (Intr)
time         82.82339 -0.446
Residual    215.70104

Fixed effects: list(my.form)
                Value Std.Error   DF    t-value p-value
(Intercept)  794.3135 26.590386 1248  29.872208  0.0000
time        -122.1458  7.453226 1248 -16.388309  0.0000
....
```

- ⋆ substantial variation in the random slopes relative to that in the random intercepts and residual error
- ⋆ negative correlation between the random intercepts and slopes



- ⋆ vast majority of subject-specific slopes are negative
- ⋆ participants with higher initial CD4 count tend to have steeper declines post-seroconversion

```
> ## Random intercepts + (exponential spatial and independent measurement error)
> ##
> summary(fit6.ML)
Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:      156.79 297.0372

Correlation Structure: Exponential spatial correlation
 Formula: ~time | id
 Parameter estimate(s):
   range   nugget
2.534943 0.386866
...
>
> ## Correlation between two observation 1 time unit apart
> ##
> (1-nugget) * exp(-1/range)
[1] 0.4132684
```

- Fairly substantial correlation in the serial dependence term

- Point estimates:

$$\text{E}[\text{CD4}_{ki}] = \beta_0 + \beta_1 \text{time}_{ki} + \beta_2 \text{age}_{ki} + \beta_3 \text{smoker}_{ki} \\ + \beta_4 \text{drug}_{ki} + \beta_5 \text{cesd-CS}_{ki} + \beta_5 \text{cesd-LO}_{ki}$$

Dependence model	Int	time	age	smoker	drug	cesd-CS	cesd-LO
0.	720	-85	2	168	58	-8	17
1.	791	-104	-1	101	19	-8	18
2.	788	-121	4	106	28	-10	16
3.	794	-122	3	92	28	-7	16
4.	918	-134	7	100	20	-22	10
5.	797	-108	1	103	23	-7	13
6.	811	-112	1	97	26	-9	14

- Standard error estimates:

$$\begin{aligned} \text{E}[\text{CD4}_{ki}] \;=\; \beta_0 \;+\; \beta_1\text{time}_{ki} \;+\; \beta_2\text{age}_{ki} \;+\; \beta_3\text{smoker}_{ki} \\ +\; \beta_4\text{drug}_{ki} \;+\; \beta_5\text{cesd-CS}_{ki} \;+\; \beta_5\text{cesd-LO}_{ki} \end{aligned}$$

Dependence model	Int	time	age	smoker	drug	cesd-CS	cesd-LO
0.	21.4	5.7	5.7	18.0	20.4	4.9	5.8
1.	26.1	4.7	10.4	23.9	21.1	8.0	5.3
2.	26.1	6.8	10.6	24.0	21.3	8.1	5.3
3.	26.6	7.5	10.3	23.9	21.0	7.9	5.2
4.	31.2	7.9	14.6	26.9	22.1	10.8	5.4
5.	26.6	5.7	10.5	24.3	21.2	8.0	5.3
6.	27.3	6.3	10.8	24.6	21.3	8.2	5.2

Model diagnostics

- Since estimation/inference for linear mixed effects models is likelihood-based, the validity of results relies on a series of assumptions:
  - correct specification of the mean model
  - correct specification of the dependence model
  - Normality of $\boldsymbol{\gamma}_k$ and $\boldsymbol{\epsilon}_k$
  - K is sufficiently 'large' for asymptotic results to apply
- Prior to formal modeling, one can (and should) perform an initial EDA to examine possible errors in the data as well as to get insight into model structure
- Once a model has been fit, the assumptions can be re-assessed using an analysis of residuals and the empirical Bayes estimates of the random effects

Residuals

- In mixed effects models, the notion of a ‘residual’ can be defined at a number of different levels
  - ⋆ **marginal** (population-level) residuals:

$$\boldsymbol{e}_k \;=\; \boldsymbol{Y}_k \;-\; \boldsymbol{X}_k\boldsymbol{\beta}$$

  - ⋆ **stage-one** (cluster-level) residuals:

$$\boldsymbol{\epsilon}_k \;=\; \boldsymbol{Y}_k \;-\; \boldsymbol{X}_k\boldsymbol{\beta} \;-\; \boldsymbol{Z}_k\boldsymbol{\gamma}_k$$

  - ⋆ γ_k can also be viewed as form of **stage two** residual
    - ∗ deviation around the population regression

- Estimates of these quantities can be obtained by plugging in $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\gamma}_k$:

$$\widehat{\boldsymbol{e}}_k \;=\; \boldsymbol{Y}_k \;-\; \boldsymbol{X}_k\widehat{\boldsymbol{\beta}}$$

$$\widehat{\boldsymbol{\epsilon}}_k \;=\; \boldsymbol{Y}_k \;-\; \boldsymbol{X}_k\widehat{\boldsymbol{\beta}} \;-\; \boldsymbol{Z}_k\widetilde{\boldsymbol{\gamma}}_k$$

Standardization

- The observed residuals arise from having fit a model and, therefore, should exhibit variation-covariation in accordance with the structure of the model
  - ⋆ e.g. if the model permits the residual error variance to differ by some covariate, then the observed residuals will reflect this

- If the goal is to investigate the extent to which the model is 'inadequate', we should acknowledge this phenomenon

- One way forward is to standardize (or normalize) the residuals
  - ⋆ let $\boldsymbol{L}_k^T \boldsymbol{L}_k$ be the Cholesky decomposition of $\boldsymbol{\Sigma}_k(\widehat{\boldsymbol{\alpha}})$
  - ⋆ use $\boldsymbol{L}_k$ to form the **standardized marginal** residuals

$$\begin{aligned} \widehat{\boldsymbol{e}}_k^* &= \boldsymbol{L}_k^{-1} \widehat{\boldsymbol{e}}_k \\ &= \boldsymbol{L}_k^{-1}(\boldsymbol{Y}_k - \boldsymbol{X}_k \widehat{\boldsymbol{\beta}}) \end{aligned}$$

- If $\boldsymbol{\Sigma}_k(\boldsymbol{\alpha})$ has been correctly specified then $\text{Cov}[\widehat{\boldsymbol{e}}_k^*] \approx \boldsymbol{I}_{n_k}$
  - ⋆ no mean-variance relationship in the variance components
  - ⋆ no relationship between the residuals and any of the covariates
  - ⋆ no residual correlation among observations within a cluster

- Note, as appropriate, one might also consider standardizing the stage one residuals

$$\boldsymbol{\epsilon}_k = \boldsymbol{Y}_k - \boldsymbol{X}_k\boldsymbol{\beta} - \boldsymbol{Z}_k\boldsymbol{\gamma}_k$$

  - ⋆ particularly useful for investigating residual (unaccounted) serial dependence in the specification of $\boldsymbol{R}_k(\boldsymbol{\alpha})$

Diagnostics

- Conventionally, folks tend to focus on the stage one residuals and the random effects estimates when performing diagnostics for linear mixed models
  - ⋆ patterns in the population residuals could result from inadequacy in the model structure in various places

- Mean model:
  - ⋆ plots of $\hat{\epsilon}_{ki}$ versus the columns of $\boldsymbol{X}_k$ to investigate whether the assumed specification of $\boldsymbol{X}_k\boldsymbol{\beta}$ is adequate
  - ⋆ plots of $\tilde{\gamma}_{kq}$ versus the columns of $\boldsymbol{X}_k$ to investigate whether the assumed specification of $\boldsymbol{Z}_k\boldsymbol{\gamma}_k$ is adequate

- Note, if the residuals have been standardized, one should compare them with the columns of $\boldsymbol{X}_k$ similarly standardized

- Dependence model:
  - ⋆ plot of $\hat{\epsilon}_{ki}$ versus the fitted values, $\hat{\mu}_{ki} = \boldsymbol{X}_k\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_k\widetilde{\gamma}_k$, to see if there is a residual mean-variance relationship
  - ⋆ plot of $\hat{\epsilon}_{ki}$ versus lag($\hat{\epsilon}_{ki}$) to see if there is any residual serial dependence

- Normality:
  - ⋆ standard approach is to present a Q-Q plot of the fitted residual versus the expected normalized residuals from a Normal distribution
    - ∗ stage one residuals, $\hat{\epsilon}_{ki}$
    - ∗ random effect estimates, $\tilde{\gamma}_{k,j}$ for $j = 1, \ldots, q$
  - ⋆ as appropriate, one can also examine scatter plots of the components of $\widetilde{\gamma}_k$ to assess joint Normality

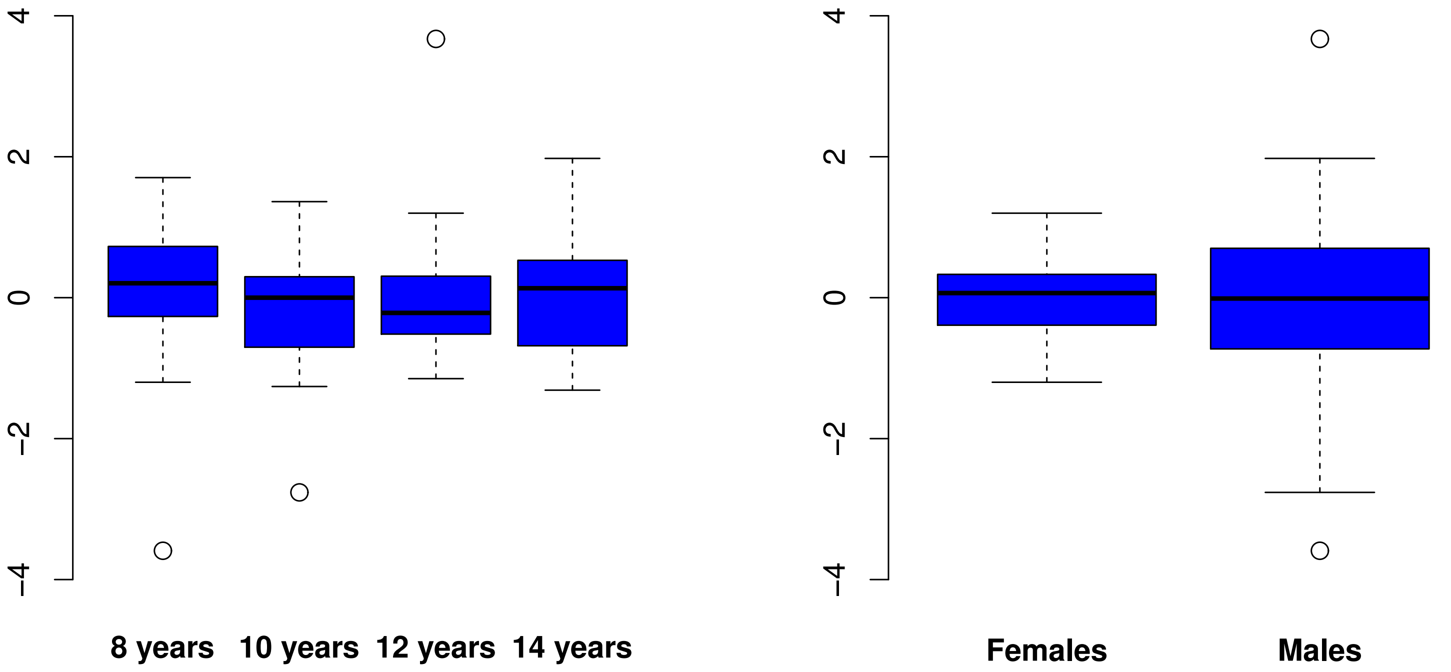
Dental growth data

- Recall the 'best' fitting model was model 8
  - ⋆ random intercepts + heteroskedastic error terms by gender

- Present select diagnostics based on model 1:
  - ⋆ random intercepts + homoskedastic independent error

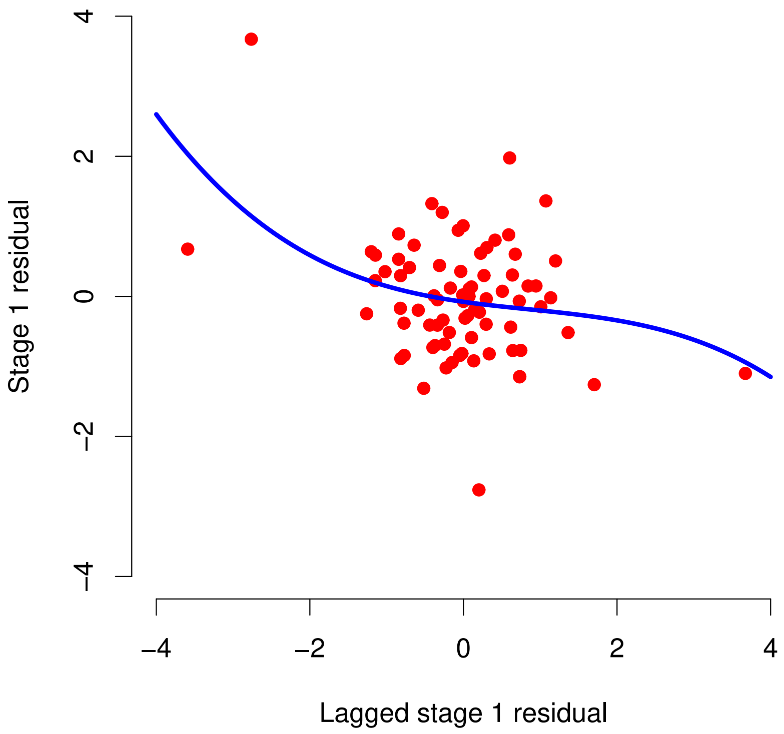
```
>
> ## Random intercepts + independent errors
> ##
> fit1.ML <- lme(fixed=length ~ ageStar * gender,
+                random=reStruct(~ 1 | id),
+                data=growth, method="ML")
>
> epsHat   <- resid(fit1.ML, type="normalized")
> gammaHat <- ranef(fit1.ML)[,1]
```

- Investigate heteroskedasticity in the standardized stage 1 residuals:



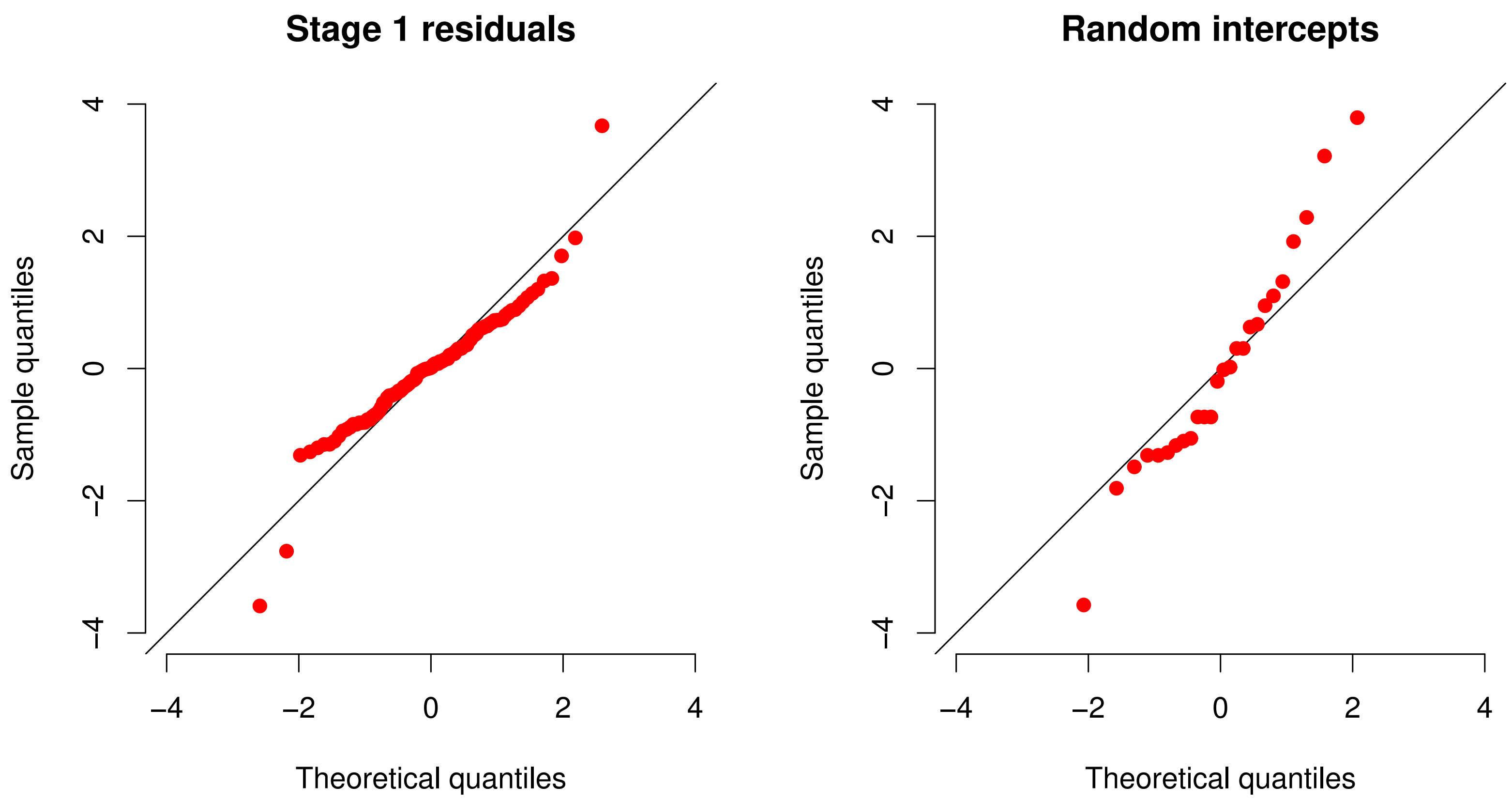
- ⋆ inconclusive evidence regarding heteroskedasticity by age
- ⋆ strong evidence of heteroskedasticity by gender

- Compare residuals and lagged residuals to investigate potential serial dependence:
  - ⋆ only for observations at ages 10, 12 and 14



  - ⋆ indication of unaccounted for local correlation

- Q-Q plots:



  - ⋆ looking for linearity between the theoretical and sample quantiles
  - ⋆ some evidence of heavier-than-Normal tails in the stage 1 residuals
  - ⋆ random intercepts seems fine
    - ∗ tough to tell with only $K=26$ children

Summary

- Goals:
  - ⋆ perform estimation/inference for regression parameters from a model for a continuous response while acknowledging within-cluster dependence
  - ⋆ also perform estimation/inference for the dependence structure
  - ⋆ estimate cluster-specific parameters/effects

- Approach:
  - ⋆ combine *fixed effects* with *random effects* into a single model formulation:

$$\boldsymbol{Y}_k = \boldsymbol{X}_k\boldsymbol{\beta} + \boldsymbol{Z}_k\boldsymbol{\gamma}_k + \boldsymbol{\epsilon}_k$$

  - ⋆ consider various strategies for structuring the model components so that a range of dependence models can be entertained
    - ∗ structure for Cov$[\boldsymbol{\gamma}_k] = \boldsymbol{G}(\boldsymbol{\alpha})$
    - ∗ structure for Cov$[\boldsymbol{\epsilon}_k] = \boldsymbol{R}_k(\boldsymbol{\alpha})$

- Estimation/inference:
  - ⋆ integrated likelihood for $(\boldsymbol{\beta}, \boldsymbol{\alpha})$
  - ⋆ ML or REML
  - ⋆ empirical Bayes for the cluster-specific random effects, $\boldsymbol{\gamma}_k$
- Since estimation/inference is fully parametric, diagnostics for model structure and assumptions are an important consideration
  - ⋆ mean model
  - ⋆ dependence model
  - ⋆ distributional assumptions