Problem n.3

The file `School.txt` contains data regarding 5000 pupils who attend 40 different primary schools. Students are tested in mathematics with a standardized test across schools. The response variable $\texttt{Achiev} \in \mathbb{R}$ is the normalized achievement test score. Three explanatory variables at student level are provided:

- $\texttt{ESCS} \in \mathbb{R}$: a scale centered in 0 for pupil socioeconomic status;
- $\texttt{Gender} \in \{0, 1\}$: 1 for male, 0 for female;
- $\texttt{Bilingual} \in \{0, 1\}$: 1 if at least one of the pupil's parent is a foreigner with respect to the country in which the pupil studies, 0 otherwise.

Moreover, $\texttt{IDSchool} \in \{1, ..., 40\}$ provides the anonymous school identification number.
Consider the following linear mixed-effects model:

$$\texttt{Achiev}_i = \beta_0\, \mathbb{1}_i + \beta_1\, \texttt{Gender}_i + \beta_2\, \texttt{ESCS}_i + \beta_3\, \texttt{Bilingual}_i + b_{0i}\, \mathbb{1}_i + b_{1i}\, \texttt{ESCS}_i + \boldsymbol{\epsilon}_i \quad \text{for} \quad i \in \texttt{IDSchool}$$

where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}_i \boldsymbol{\mathcal{C}}_i \boldsymbol{\Lambda}_i)$ and

$$\boldsymbol{b}_i = \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \right)$$

a) Assume homoscedastic residuals and $d_{12} = d_{21} = 0$. Fit the model (**M1**) and compute d_{11}, d_{22} and σ^2.

b) Compute the PVRE for **M1**.

c) Fit now a model (**M2**) with the same fixed part of **M1**, but now *without the random slope* and with *uncorrelated heteroscedastic* residuals with

$$\boldsymbol{\Lambda}_i = \begin{bmatrix} \lambda_1^{(i)} & 0 & .. & 0 \\ 0 & \lambda_2^{(i)} & .. & 0 \\ .. & .. & .. & 0 \\ 0 & 0 & .. & \lambda_{125}^{(i)} \end{bmatrix}$$

and $\boldsymbol{\lambda}^{(i)} = [\lambda_1^{(i)} \quad \lambda_2^{(i)} \quad ... \quad \lambda_{125}^{(i)}]' = \sqrt{\texttt{ESCS}_i}$ ($\boldsymbol{\lambda}_i$'s known case), for $i \in \texttt{IDSchool}$. Compute d_{11} and σ^2.

d) Report the dot plot of the estimated random intercepts in **M2**. Net of the impact of fixed effect covariates, which is the `IDSchool` associated to the lowest achievement?

Upload your solution **here**