**Politecnico di Milano**
**Scuola di Ingegneria Industriale e dell'Informazione**
Applied Statistics
Exam 2025-06-12 - Part B - 2024/2025

# Problem 1: Modelling the impact of indoor greenery on mood in high-rise living

The file `plants_mood.txt` contains data from 2000 residents of high-rise apartment buildings across 20 different urban zones in Southeast Asia, identified by the variable `zone_id`. Participants took part in a behavioral ecology project called *Green Within Walls*, which examined how indoor plant ownership affects daily well-being in dense urban environments. Each resident logged daily mood scores via an app and completed surveys on their indoor environments.

Consider the following linear mixed-effects model:

$$\underline{\texttt{mood}}_i = \beta_0 \underline{1}_{n_i} + \beta_1 \underline{\texttt{plants}}_i + \beta_2 \underline{\texttt{light}}_i + \beta_3 \underline{\texttt{social}}_i + b_i \underline{1}_{n_i} + \underline{\epsilon}_i \qquad \textbf{(M1)}$$

for $i \in$ `zone_id`, with $\underline{\epsilon}_i \sim \mathcal{N}(\underline{0}, \sigma^2 I_{n_i})$, $b_i \sim \mathcal{N}(0, \sigma_b^2)$, and $n_i$ the number of participants in zone $i$.

- $\underline{\texttt{mood}}_i$ is the $n_i$-dimensional vector of **average daily mood scores** (mood score is on a 0–10 scale) for individuals in urban zone $i$;
- $\underline{\texttt{plants}}_i$ is the $n_i$-dimensional vector representing the **number of indoor plants** owned per person;
- $\underline{\texttt{light}}_i$ is the $n_i$-dimensional vector indicating **average hours of natural light exposure** inside the apartment per day;
- $\underline{\texttt{social}}_i$ is the $n_i$-dimensional vector of **average hours of in-person interaction** per day (e.g., roommates, neighbors, community spaces).

a) Fit model **M1**, estimate $\beta_1$. Is there a significant effect of the number of plants on the well-being? Compute the PVRE.

b) Fit an extended model **M2**, allowing heteroscedastic residuals: $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$ with

$$\sigma_{ij} = \sigma \cdot |\texttt{social}_{ij}|^{\delta}$$

for individual $j \in \{1, \ldots, n_i\}$, in zone $i \in$ `zone_id`.

Estimate $\delta$ for model **M2**.

c) Should **M2** be preferred over **M1**? Support your answer with a test.

d) Estimate (using **M2**) the mood of a person having 5 plants, 12 hours of natural light exposure and 5 hours of in-person interaction per day.

Upload your results here: `https://forms.cloud.microsoft/e/FvMhh6nqTX`