# Problem n.1

The file `ElectronicSales.txt` contains data regarding 30 electronic stores, identified with the variable `store_id`, located in Switzerland. For each store, we have data on:

- Whether it is located within a metropolitan area (`metro_area` $\in \{0, 1\}$);

- The total floor area of the store in $m^2$ (`size` $\in \mathbb{R}^+$);

- The quarterly sales revenue for 2023 (identified with the variables `sales` $\in \mathbb{R}^+$ and `quarter` $\in \{1, 2, 3, 4\}$).

Additionally, we have baseline information for the fourth quarter of 2022, including the initial sales revenue (`sales0` $\in \mathbb{R}^+$) and a measure of customer satisfaction (`satisfaction` $\in \{0, 1\}$).

a) Implement the following linear regression model **M0**:

$$\texttt{sales} = \beta_0 + \beta_{1,q} + \beta_2\,\texttt{sales0} + \beta_{3,q}\,\texttt{satisfaction} + \beta_4\,\texttt{metro\_area} + \beta_5\,\texttt{size} + \epsilon\,, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.,} \quad \textbf{(M0)}$$

for every quarter $q \in \{1, 2, 3, 4\}$.

Briefly detail the implementation choices for **M0** reporting also the relevant R code. Report the estimates of $\beta_0$ and $\beta_5$, the standard deviation $\sigma$ of the error term and the AIC.

b) Analyze the standardized residuals. Plot the residuals versus the fitted values and comment the plot. Visualize the conditional distribution of the residuals given `quarter` and comment it.

In your opinion, what factors or aspects of the problem does **M0** fail to consider?

c) Implement a model **M1** in which the error term for store $i$ and quarter $q$ is $\epsilon_{i,q} \sim \mathcal{N}(0, \sigma^2 |\mu_{i,q}|^{2\delta})$ distributed where $\mu_{i,q}$ is the mean value of `sales`.

Briefly detail the implementation choices for **M1** reporting also the relevant R code. Report the estimates of $\beta_0$, $\delta$, $\sigma$ and the AIC of **M1**.

d) What is the criticality when estimating the parameters in this group?

e) Implement now a model **M2** with *VarPower()* variance function in the $< \delta >$-group, stratifying by `metro_area` (assume $\boldsymbol{\delta} = [\delta_0, \delta_1]$ with $\delta_0$ if `metro_area` $= 0$ and $\delta_1$ if `metro_area` $= 1$).

Briefly detail the implementation choices for **M2** reporting also the relevant R code. Report $\delta_0$ and $\delta_1$; interpret their values. Is there statistical evidence to assume $\delta_0 \neq \delta_1$?

f) Implement a model **M3** with the same within-group heteroscedasticity structure of **M2**, but with general correlation matrix for residual errors. Estimate the unknown parameters in the correlation matrix, reporting also the relevant R code, and the standard deviation $\sigma$ of the error term.

What is the best model among **M1**, **M2** and **M3**?

Upload your solution [https://forms.office.com/e/qhvhfeM41W](https://forms.office.com/e/qhvhfeM41W)