

Problem n.2

The file `Urban_property_values.txt` contains data for the year 2023 regarding the 107 provinces in Italy (`ID_area`) grouped by their regional code (`ID_reg`). For each province, the observed values of the following variables are provided:

- the public transport index (`transport_index` $\in \mathbb{R}$),
- the average number of cultural events (`avg_cultural_events` $\in \mathbb{R}$),
- the average air quality index (`avg_air_quality` $\in \mathbb{R}$),
- the average household income (`avg_household_income` $\in \mathbb{R}$),
- the average crime rate (`avg_crime_rate` $\in \mathbb{R}$),
- the average property value (`avg_property_value` $\in \mathbb{R}$).

Moreover, the values of the 2-level variable `gentrified` $\in \{\text{Yes, No}\}$ are also provided, indicating whether a province is gentrified or not (i.e., whether it changed from being a poor area to a richer one, by people of a higher social class moving to live there).

- a) Formulate a classical linear regression model (**M0**) for `avg_property_value` as a function of all the other continuous variables and the categorical variable `gentrified`, with no interactions.
 1. Briefly detail its implementation reporting also the relevant R code.
 2. Report the standard deviation σ of the error term $\epsilon \sim \mathcal{N}(0, \sigma^2)$, together with the adjusted R-squared and the AIC.
- b) Does the average number of cultural events have a positive or negative effect on average property value? Is that significant?
- c) Update now the model **M0**, introducing a compound-Symmetry Correlation Structure and using the region as grouping factor (model **M1**).
 1. Briefly detail its implementation reporting also the relevant R code.
 2. Report the values of the intercept and the AIC.
 3. Compute and report (in the form $[\min, \max]$) the 95% confidence intervals for ρ and σ .
 4. Comment on the obtained values and draw your conclusions.
- d) Fit a mixed-effects model with random intercept (model **M2**) starting from model **M0** and introducing a random intercept related to the regional grouping factor.
 1. Briefly detail its implementation reporting also the relevant R code.
 2. Report the AIC and the standard deviations of the random intercept and the error term.
 3. Compare **M1** and **M2** and draw your conclusions.

Upload your solution <https://forms.office.com/e/W3L9VC8U2y>