

## Problem 3: Employee's monthly expenditure

The file `expenditure.txt` provides information on the financial habits of 200 employees in a company. Specifically, the following variables are available for each employee:

- `avg_exp` (numeric): average monthly credit card expenditure in euros
- `income` (numeric): gross annual income in thousands of euros
- `age` (int): age of the employee in years
- `perc_taxes` (numeric): percentage of salary spent on taxes
- `owns_house` (int): indicates whether the employee owns or rents the residence (0 = owns, 1 = rents)

We consider the following linear model for explaining `avg_exp`:

$$\text{avg\_exp}_i = \beta_0 + \beta_1 \text{income}_i + \beta_2 \text{age}_i + \beta_3 \text{perc\_taxes}_i + \beta_4 \text{owns\_house}_i + \epsilon_i \quad (1)$$

for  $i \in \{1, \dots, n\}$  with  $\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$ ,  $n = 200$

- Fit model [1](#) using Ordinary Least Squares and report the estimates of the unknown parameters.
- Do the assumptions for the residuals hold? By inspecting the distribution of the residuals against each variable, propose a new model which might include an appropriate additional term. Are the assumptions for the residuals recovered for the new model?
- What is the mean difference in average monthly expenses between employees who own a house and those who rent? Provide a 95% confidence interval.
- One could approach the fitting of model [1](#) (without the modification brought in b)) using a penalized regression method. What are the motivations for considering a penalized regression in this case? Provide a precise justification for your answer.
- For this question, the random state should be fixed by inserting `set.seed(20230707)` at the beginning of the code.*

Fit model [1](#) with Lasso regression, using a sequence `lambda = 10^seq(10, -2, length.out=100)`. Considering the 10-fold cross-validation error but also the model complexity, which value of `lambda` would you choose?

Upload your results here:

<https://forms.office.com/e/ZwOKYsivCk>