

## Problem n.3

We aim to explore the determinants that shape employee productivity within a software development company that operates in 20 separate and decentralized locations. The dataset `productivity.txt` encompasses productivity ratings of 400 employees, together with their years of `experience`, `education` level (ranging among `HighSchool`, `Degree`, `PhD`), average `weekly` working hours, and designated office `location`. Our objective is to analyze the impact of these variables on employee productivity through a linear model with the following structure:

$$\text{productivity} = \beta_0 + \beta_1 \text{experience} + \beta_2 \text{education} + \beta_3 \text{hours} + \epsilon, \quad \text{with } \epsilon \sim \mathcal{N}(0, \sigma^2).$$

- a) Fit the model and provide estimates for the unknown parameters.
- b) Check the assumptions of the model.
- c) Conduct a hypothesis test at the 5% significance level to determine if `experience` significantly affects employee `productivity`. What is the increment/decrement of the `productivity` of an employee with 10 years of `experience`?
- d) Provide an interpretation of the impact of the `education` on the `productivity`.
- e) Now consider a modified model where the variable `location` is added as a random intercept. Fit this new model and report the PVRE.
- f) Create a dot plot displaying the estimated random intercepts. Disregarding the fixed effect covariates, which `location` is associated with the highest productivity score?

Upload your solution [here](#)