Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
Exam 2024-06-13 - Part B - 2023/2024

Problem 2: Effect of temperature on tomatoes growth rate

We are studying the effect of temperature on the growth rate of tomatoes. The file `tomatoes.txt` contains data regarding 2089 greenhouses growing 200 different species of tomatoes. Specifically it reports:

- the average temperature in the greenhouse during the growth season: `temp` (the variable is centred)
- the tomato species that is cultivated, encoded as an integer variable: `species`
- the average yield per crop cycle in kg/m^2: `yield`

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

$$\texttt{yield} = \beta_{0,i} + \beta_1 \, \texttt{temp} + \epsilon \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$, and $i \in \{1, \dots, 200\}$ indicates the tomato species.

Report the estimates of $\beta_{0,2}$ and β_1.

b) Consider now the model **M1**:

$$\texttt{yield}_i = \beta_0 + b_i \mathbf{1}_{n_i} + \beta_1 \, \texttt{temp}_i + \boldsymbol{\epsilon}_i \tag{2}$$

with $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_{n_i})$, $b_i \sim \mathcal{N}(0, \sigma_b^2)$ and n_i the number of greenhouses cultivating species i.

Fit the model and report the estimate of σ_b. Without performing any model comparison, in your opinion, what is the advantage of **M1** over **M0**?

c) A farmer tells you: "High temperatures generally favour tomatoes growth but this effect is more or less pronounced depending on the species that is considered". Propose and fit an update **M2** of **M1** to account for the effect described by the farmer. Is there a species for which we estimate that the temperature has a negative effect?

d) Comment on whether **M1** or **M2** is better, supporting your answer with a test.

Upload your results here: `https://forms.office.com/e/RgMXqMMfyg`