

Problem n.2

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

- the average temperature in the field during the growth season: `temp` (the variable is centred)
- the wine grape species that is cultivated, encoded as an integer variable: `species`
- the average yield per crop cycle in kg/m²: `yield`

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

$$\text{yield} = \beta_{0,i} + \beta_1 \text{temp} + \epsilon \quad (1)$$

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

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

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

$$\underline{\text{yield}}_i = \beta_0 \mathbb{I}_{n_i} + b_i \mathbb{I}_{n_i} + \beta_1 \underline{\text{temp}}_i + \underline{\epsilon}_i \quad (2)$$

with $\underline{\epsilon}_i \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_{n_i})$, $b_i \sim \mathcal{N}(0, \sigma^2 d_{11})$, n_i the number of fields cultivating species i and \mathbb{I}_{n_i} the identity matrix of size $n_i \times n_i$. Moreover, $\underline{\text{yield}}_i$ and $\underline{\text{temp}}_i$ are the vectors that contain the yield and temperature data, respectively, for all fields cultivating species i .

Fit the model and report the estimate of σ and d_{11} . 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 wine grape 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 solution <https://forms.office.com/e/eUD4uHqFrD>