

## Problem n.3

The site WineReviews collects the reviews of many bottles of wine from across the world. The file `wineReviews.txt` contains the points obtained (the higher the value, the better), together with the price and the percentage of alcohol for 547 bottles from Italy. We are interested in studying if the price and the percentage of alcohol influence the number of points via a linear model. Consider the following linear models:

$$\text{points} = \beta_0 + \beta_1 \cdot \text{price} + \beta_2 \cdot \text{alcohol} + \varepsilon \quad (1)$$

$$\ln(\text{points}) = \beta_0 + \beta_1 \cdot \ln(\text{price}) + \beta_2 \cdot \text{alcohol} + \varepsilon \quad (2)$$

with  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ .

- a) Choose between Model (1) and Model (2) the one that better describes the relationship between the variables, considering especially the diagnostics of the models.

From now on, answer to the following questions based only on the chosen model.

- b) Provide an estimate of the  $\beta_i$ ,  $i = 0, \dots, 2$ , and of  $\sigma$ .
- c) Perform a test of level 5% to verify if price and alcohol can be both discarded from the model.
- d) Perform any other statistical test that you consider useful to reduce the model, and update the estimates of its parameters.
- e) Let's consider now the variable region in the model as a random intercept. Fit a suitable model for accounting the hierarchy and compute and report the PVRE index.
- f) Report the dot plot of the estimated random intercepts. Discounting the effect of fixed effect covariates, which is the province associated to the lowest number of points?

Upload your results here:

<https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9ASOGWf21HjvGX24HiqFVUNTmZVVFURTBNS1JFU1QxTlM4OENUTjhITi4u>