

Problem n.3

We are interested in studying the danceability (Y) of a song with respect to other features of the song. The file `danceability.txt` contains the values of danceability (the higher the value, the easier it is to dance to this song), loudness, energy, tempo and genre of 400 songs. Consider a linear model of the form:

$$Y = \beta_0 + \beta_1 \cdot \text{loudness} + \beta_2 \cdot \text{energy} + \beta_3 \cdot \text{tempo} + \varepsilon$$

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

- a) Provide an estimate of the β_i , $i = 0, \dots, 3$, and of σ .
- b) State and verify the model assumptions.
- c) Perform a test of level 5% to verify if loudness and energy can be both discarded from the model.
- d) Perform any other statistical tests that you consider useful to reduce the model, and update the estimates of its parameters.
- e) Let's consider now the variable genre 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. Net to the effect of fixed effect covariates, which is the genre associated to the highest danceability?

Upload your results here:

<https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9ASOGWf2lHjvGX24HiqFVURTNSSEk0WktLNU5PRTJXOVM10ENKS1Y0Ty4u>