**Politecnico di Milano**
**Scuola di Ingegneria Industriale e dell'Informazione**
Applied Statistics
February 9th, 2022

# Problem n.3

The file `wine.txt` reports the data on the alcohol content in 179 bottles of wine. For the alcohol content consider a linear model, accounting for the sugar content of grapes, and for type of wine ('Red', 'Rose', 'White'):

$$\text{alcohol}_g = \beta_{0,g} + \beta_{1,g} \cdot \text{sugar} + \epsilon,$$

with $\epsilon \sim N(0, \sigma^2)$ and $g$ the grouping structure induced by the type of wine.

a) Estimate the parameters of the model ($\{\beta_{0,g}, \beta_{1,g}, \sigma\}$). Verify the model assumptions, reporting any plot you consider important.

b) Perform two statistical tests – each at level 1% – to verify if

- there is a significant dependence of the mean alcohol content on the type of wine;
- there is a significant dependence of the mean alcohol content on the sugar content.

c) Based on tests (b) or any other test deemed relevant, reduce the model and report the updated model parameters.

d) Build a prediction interval at 99% for a new bottle of red wine made with grapes with 20 g of sugar.

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9AS0GWf2lHjvGX24HiqFVUMlFYTFZJWTE1QkZUOFROU0xJODg5T1lLOS4u`