**Politecnico di Milano**
**Scuola di Ingegneria Industriale e dell'Informazione**
Applied Statistics
September 5th, 2023



# Problem n.1

A study is conducted to analyze the impact of weather conditions and shop position on the sales of ice cream from June to August 2023 in Milan. The dataset `icecreams.txt` includes the UTM geographical coordinates $s_i$ of various ice cream parlours, the recorded average daily temperature $\text{temp}_i$ (in degrees Celsius) during this period and the daily sales $y(s_i)$ [k€/day]. The latters are modeled based on temperature using the following relationship:

$$y(s_i) = b_0 + b_1 \ \text{temp}_i + \delta(s_i), \tag{1}$$

where $\delta(s_i)$ represents stationary residual with *spherical variogram without nugget*.

a) Estimate the parameters $b_0$ and $b_1$ using the generalized least squares method. Discuss the model assumptions.

b) Report the fitted variogram and the related estimated values.

c) Compute the average sales prediction for the month of July 2024 for an ice cream parlour with an average temperature of 30°C.

d) Due to the geographical location of the parlours, they can be categorized into central (`central`=1) or peripheral (`central`=0). Modify the model in Eq. (1) to include this categorical effect, as follows

$$y(s_i) = b_{0,j} + \delta(s_i) \tag{2}$$

where $j$ is the grouping induced by the variable `central`. Estimate the parameters and interpret the coefficients.

e) Which model do you deem more appropriate to describe the data? Comment on your choice.

**Upload your solution here**