

Problem 4: Spatial modelling of the sales of cold beverages

A study is conducted to analyze the impact of weather conditions and geographic location on the sales of cold beverages by bars from July to September 2023 in Milan. The dataset `beverages.txt` includes the UTM geographical coordinates s_i of various bars, the recorded average daily temperature temp_i (in degrees Celsius) during this period and the daily sales of cold beverages $y(s_i)$ [k€/day]. Consider the following model:

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

where $\delta(s_i)$ represents stationary residuals 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 prediction of the total sales for the month of July 2024 for a bar with an average temperature of 30°C.
- d) Due to the geographical location of the bars, 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:

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

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.