Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
January 20th, 2023

©All rights reserved. Note: data may be simulated and should not be considered outside of the scope of the present exam.

AppliedStatistics

Problem n.1

Airborne particulate matter (PM) is a complex mixture of solids and aerosols composed of small droplets of liquid, dry solid fragments, and solid cores with liquid coatings. Particles are defined by their diameter for air quality regulatory purposes: those with a diameter of 10 microns or less (PM10) are inhalable into the lungs and can induce adverse health effects. The mayor of the Comune di Milano is worried for the health conditions of his citizens and asks *AppliedStatistics*, a company selling data science analysis based in Milan, to model the distribution of PM10 in the city of Milan.

He provides the file `Pollution.txt` containing the measurements of daily maximum concentration z of `PM10` $[\mu g/m^3]$ recorded on March 18th, 2022 (the most polluted day in 2022) within 180 locations in the city of Milan. The s_k $(k = 1, ..., 180)$ locations are expressed in UTM coordinates (variables `x` and `y`) and, through the categorical variable `Area`, the indication of whether the measurement location is inside a *garden or park* (`Area` $= 0$), inside a *pedestrian area* (`Area` $= 1$) or *outside both* (`Area` $= 2$) is provided.

a) Consider for the `PM10` $z(s_k)$ the following model

$$z(s_k) = \beta_0 + \delta(s_k) \quad \text{for} \quad k = 1, ..., 180$$

where $\delta(s_k)$ a stationary residual with *Exponential* model *without Nugget* for spatial dependence structure.

Estimate the parameters β_0 and $\delta(s_k)$ of the model fitted via *Generalized Least Squares*.

b) Consider now for the `PM10` $z_i(s_k)$ the following model

$$z_i(s_k) = \beta_0 + \beta_1 \, \texttt{Area}(s_k) + \delta(s_k) \quad \text{for} \quad i = 1, 2, 3 \;\; k = 1, ..., 180$$

where $\delta(s_k)$ a stationary residual with *Spherical* model *without Nugget* for spatial dependence structure.

Estimate the parameters β_0 and β_1 of the model fitted via *Generalized Least Squares*.

c) Use the model at point (b) to provide a point prediction $z^*(s_0)$ for the maximum concentration of `PM10` on March 18th, 2022 at location s0 = (514961.61, 5034538.28) (within the Duomo square, therefore a pedestrian area).

d) By using the same model, estimate now the maximum concentration of `PM10` in Piazzale Loreto, one of the busiest squares of Milan, on March 18th, 2023.

Upload your solution **here**