Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
January 25th, 2024

Problem n.2

The manager of the hotel *Boule de Neige* in the Italian Alps is designing the pricing strategy for 2024. He knows that there are generally three types of periods (*low* demand (L), *medium* demand (M), or *high* demand (H)) and he wants to use this information to define his prices.

In the file `prices.txt` the historical series of the average price per night y [€/night] for the hotels in the local area where *Boule de Neige* is located collected in 2023. Assuming that the prices are distributed according to a $\mathcal{N}(\mu_L, \sigma_L^2)$ when the demand is *low*, $\mathcal{N}(\mu_M, \sigma_H^2)$ when the demand is *medium* and $\mathcal{N}(\mu_H, \sigma_H^2)$ when the demand is *high*.

Remark: before answering the following questions set *set.seed(1)*.

a) Estimate μ_L, σ_L, μ_M, σ_M, μ_H and σ_H and comment on the model used to reach these estimates.

b) How could the parameters estimated at point a) be used to set up the pricing strategy in 2024?

c) Compute the transition probabilities $p_{(L,M)}$ and $p_{(L,H)}$ of moving from a *low* demand periods, to a *medium* and a *high* demand periods, respectively.

The manager of the neighbouring *White Winter* hotel wants to use a model for the definition of his pricing strategy, based on the spatial distribution of the neighbouring hotels prices. Specifically, he collects in the file `hotels.txt` the prices per night `y` [€/night] in 55 neighbouring hotels, observed on the Feb. 25th 2023. The dataset also reports the UTM coordinates s_i of the hotels, whether the price refers to a day during the winter season or not (`winter` $= 1$ for *yes* or `winter` $= 0$ for *no*), and the distance of the considered hotel from the funicolar connecting to the ski slopes, with $\texttt{d}(s_i) = ||s_i - s_f||$,with s_f =(342362.58, 5072518.24).

Consider for the price the following model:

$$\texttt{y}(s_i) = a_0 + a_1\,\texttt{winter} + a_2\,\texttt{d}(s_i) + a_3\,\texttt{winter}:\texttt{d}(s_i) + \delta(s_i), \tag{1}$$

with $\delta(s_i)$ a stationary residual with a spherical variogram with nugget.

d) Assuming $a_2 = a_3 = 0$, estimate the parameters a_0 and a_1 of the model via *generalized least squares*. Report the sill, the range; discuss the model assumptions.

e) Now, for general a_2 and a_3, estimate the parameters a_0, a_1, a_2, a_3 of the model via *generalized least squares*. Explain the procedure and report the point estimates of a_0, a_1, a_2 and a_3.

f) Suggest to the hotel manager a pricing strategy for a stay of 2 nights at *White Winter* (s_0) in the period Feb. 17 th to Feb. 18th 2024 (`winter` $= 1$, $s_0 = (342399.74, 5072272.75)$). Motivate your response and detail your assumptions.

Remark: remember that $\texttt{d}(s_i) = ||s_0 - s_f|| = \sqrt{(s_{0,x} - s_{f,x})^2 + (s_{0,y} - s_{f,y})^2}$

Upload your solution here **https://forms.office.com/e/D4cH8DsLEW**