Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
September 5, 2024

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

Problem n.1

A study is conducted to analyze the spatial distribution of copper concentration in the Andes Mountains in Chile. The dataset `copper.txt` reports the copper concentration at various locations within the study area. It includes the UTM geographical coordinates s of the sampling locations, a categorical variable `rock_type` indicating the type of rock (granite, sandstone, or shale), the depth of the sample `depth` (in meters), and the measured copper concentration $y(s)$ [ppm]. Consider the following model:

$$y(s) = b_{0,j} + b_1 \texttt{depth}(s) + \delta(s)$$

where $\delta(s)$ represents a stationary residual with a spherical variogram with a nugget effect and $j = 0, 1, 2$ is the grouping induced by the variable `rock_type` ($j = 0$ for granite, $j = 1$ for sandstone, $j = 2$ for shale).

a) Report a plot of the fitted variogram, *initialising* the variogram fit with the model `vgm(200000, "Sph", 2000, 100000)`. Indicate the estimate of the range and the sill.

b) Provide an estimate of the mean copper concentration at the surface of a sandstone rock type area.

c) Independently of the position, by which quantity the copper concentration increases when the sample is taken 1m lower?

d) Consider now the model update:

$$y(s) = b_{0,j} + b_{1,j} \texttt{depth}(s) + \delta(s)$$

Indicate the estimate of the sill, fitting the variogram with the same initialisation as in a). Should this model be preferred to the first one? Justify your answer.

Upload your solution https://forms.office.com/e/16Yj7864em