Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
February 7th, 2023

Problem n.3

We are interested in analysing the price of rent of beach volley courts in Milan. The file `beachVolley.txt` reports data on different rent costs as reported by the users. The dataset lists the price of the court rental [€], the length of the rental [hours], the distance from the center [Km], the distance from the nearest parking [meters], the number of available courts in the structure, the presence of shower (yes/no), the environment (indoor/outdoor), the sand color (white/red).

a) Formulate a linear regression model for the price of the court rental, as a function of all the other variables. Include in the model a possible dependence of the price on the categorical variables, but only in the intercept. Report the model and its parametrization.

b) Report and verify the model assumptions.

c) Using the appropriate statistical test, state if you deem necessary to include in the model the variable distance from the center togehter with the variable distance from the nearest parking.

d) Based on appropriate test(s), reduce the model and report the model parameters, including the estimated standard deviation.

e) You want to rent a court for 2 hours with the following characteristics: distance from the center of 1 Km, distance from the nearest parking 100 meters, 3 available courts, shower available, indoor and with white sand. Using the chosen model, compute a pointwise estimate and a prediction interval of level 95% for the price of the court rental.

Upload your results here:
`https://forms.office.com/Pages/ResponsePage.aspx?id=K3EXCvNtXUKAjjCd8ope6-9ASOGWf2lHjvGX24HiqFVUQ0g4TVVaT0pLUFhIQU4wRVpWTEgyV0JURC4u`