Factors associated with valuation of land in Colombo district: a case study

B. J. R. I. Mendis and S. D. Viswakula*

Department of Statistics, University of Colombo, Sri Lanka
*sam@stat.cmb.ac.lk

This study aims to identify the factors associated with the valuation of lands in Colombo District, and to build a model to predict the price of a land in Colombo District. Moreover, this study gives an insight on how land prices vary among the Divisional Secretariats of Colombo District. The data were collected through a telephone survey and stratified sampling technique was used treating the Divisional Secretariats as strata. The preliminary analysis of the data was carried out using appropriate data visualisations and statistical association tests. Water and electricity availability and value of additional resources that belongs to the land showed associations with price per perch according to preliminary analysis. Furthermore, it was concluded that the Colombo Division had the highest average price per perch while land values decreased as the divisions move away from the Colombo Division. Notably, Hanwella and Padukka Divisions reported the lowest price per perch. A generalised linear model with gamma distribution was fitted along with the forward selection method to select the main factors associated with the land price. A multiple linear regression model with ridge and lasso regularisations were also tried out in obtaining a better predictive model. The lasso regression model which selects features through shrinking coefficients to zero, resulted in the lowest mean squared error on the test set, in comparison to rest of the models fitted. Hence, the lasso regression model was selected as the best model for predicting the price of a land in Colombo district. The final model included the variables, Division to which the land belongs, Size of the land (in perches), Distance to nearest town, Distance to nearest bus halt, Water and electricity availability, Number of neighbouring lands to be sold, Level of the land, Type of the additional resource that belongs to the land, Value of the additional resource and the Type of land usage. The coefficient of determination and mean squared error of the proposed predictive model for the test set were 85.3% and 0.52 respectively.

Keywords: Land valuation, Colombo district, Stratified sampling, Ridge regression, Lasso regression