Goldschmidt 2023 Abstract
https://doi.org/10.7185/gold2023.19561

Statistical treatment of missing information in mineral exploration data: an application to undiscovered mineral resource assessments

MARTINA BERTELLI AND LEONARDO FELTRIN

Geological Survey of Finland

Presenting Author: martina.bertelli@gtk.fi

Data analysis is a process performed across many fields of research and businesses with the aim of extracting information and developing predictive models supporting decision-making. A common issue affecting predictivity is data missingness, which is represented by data gaps. This problem also affects data used in undiscovered mineral resource assessments, with incompleteness occurring in records of grade and tonnage of ore deposits. Several approaches have been proposed to address missing information (MI), from simple methods that disregard incomplete cases eliminating data rows (i.e., listwise deletion) to more advanced statistical techniques (maximum likelihood and multiple imputation). The selection of the most appropriate MI methodology depends on the mechanism causing data missingness and other factors [e.g., mineral exploration data records are frequently missing not at random (MNAR)]. An evaluation of five imputation algorithms is proposed, using sensitivity analysis by producing numerous synthetic regression models with variable sampling rates, known MI mechanisms, and variable proportions and distributions of MI across three variables. Multiple synthetic datasets were generated to test these methods from a public dataset on Au deposits sited in the Chugach region located in Alaska [1]. Some of the results obtained suggest that imputation algorithms in general can perform well on mineral exploration data, even if sampling rates are relatively low (small number of deposits) as comparable performances for different sampling rates (n = 25, 50, 100) were obtained. Additionally, the evaluation of model performances appears to favor either stochastic regression or not-at-random fully conditional specification (NARFCS) in both 100 and 1000 simulation rounds. Predictive mean matching algorithm appeared to perform poorly when MI was limited to a single variable if compared to other models. Although the results support usage of NARFCS, implementing this methodology may introduce additional bias connected with user interpretation of MNAR distributions.

[1] Bliss (2004), Grade and tonnage model of Chugach-type low-sulfide Au-quartz veins. In *Developments in Mineral Deposit Modeling: U.S. Geological Survey Bulletin* (pp. 44–46).