Comparative analysis of correlation coefficients in mineralogical and geophysical data from the mine tailing site “Rudnik” (Serbia)

Vesna Cvetkov, Filip Arnaut, Dragana Životić

COMPARATIVE ANALYSIS OF CORRELATION COEFFICIENTS IN MINERALOGICAL AND GEOPHYSICAL DATA FROM THE MINE TAILING SITE "RUDNIK" (SERBIA)

[1]Vesna Cvetkov, [2]Filip Arnaut, [1]Dragana Životić[1]

[1]Faculty of Mining and Geology, University of Belgrade, Djusina 7, Belgrade, Republic of Serbia
vesna.cvetkov@rgf.bg.ac.rs; dragana.zivotic@rgf.bg.ac.rs
[2]Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, Belgrade, Republic of Serbia
filip.arnaut@ipb.ac.rs

Key Words: Outlier detection, Non-linearity, Data analysis, Regression analysis, Magnetic susceptibility

INTRODUCTION

Quantifying the relationship between data is a crucial undertaking that can result in misinterpretation if the methods are not employed correctly. This communication highlights the disparity between the Pearson's correlation coefficient (PCC) and the Spearman's correlation coefficient (SCC) when used to assess the correlation between mineralogical, chemical, and geophysical data.

Although the PCC is readily accessible in commercially available software, it can yield inaccurate correlation estimates under certain assumptions. However, SCC is less commonly used and familiar to researchers in different fields. Nevertheless, it does not rely on the same assumptions as PCC, which makes it less prone to inaccurate correlation estimates. Therefore, it was considered advantageous to evaluate both correlation coefficients using a wide variety of data obtained from borehole measurements.

The research data was collected from the Rudnik mine tailing site, situated at Rudnik mountain in the Republic of Serbia. The dataset comprises mineralogical, chemical, and geophysical data pertaining to the samples collected from the boreholes.

METHODOLOGY

The purpose of a correlation coefficient, whether it is PCC, SCC, or any other type, is to assess the association between two or more variables. The assessment of the nature and magnitude of the relationship between two variables has been found to be a challenging endeavor, and the interpretation of these correlation coefficients has been primarily based on rules-of-thumb.

The PCC is a parametric measure of correlation that assumes the tested data follows a normal distribution and that the relationship between the data is linear. Microsoft Excel provides the CORREL function, which allows users to calculate PCC. Despite its availability in MS Excel, there are certain drawbacks associated with it, particularly with normally distributed data and the type of correlation between the data as not all data is normally distributed especially in natural sciences.

Conversely, SCC is a non-parametric measure of correlation that does not rely on assumptions about the distribution of the data or the nature of the correlation. SCC can be found in commercially available statistical software like JASP, but it is not included in MS Excel.

RESULTS AND DISCUSSION

The PCC and SCC correlation matrices, as well as the correlation matrix showing the difference between the PCC and SCC correlation coefficients (Figure 1), were calculated using mineralogical, chemical, and geophysical data. The most significant difference can be observed for arsenic and aluminum oxide, with a discrepancy of 0.84. Under those circumstances, the PCC was computed as 0.63, indicating a moderate correlation based on the scale provided by Schober et al. in 2018. However, the SCC was determined to be -0.21, indicating a significant discrepancy and a reversal in the correlation direction. Unlike the positive PCC, which implies that as one variable increases, so does the other, in this case, as one variable increases, the other decreases. The strength of the correlation in the SCC was determined to be weak, as per the scale provided in Schober et al. 2018.

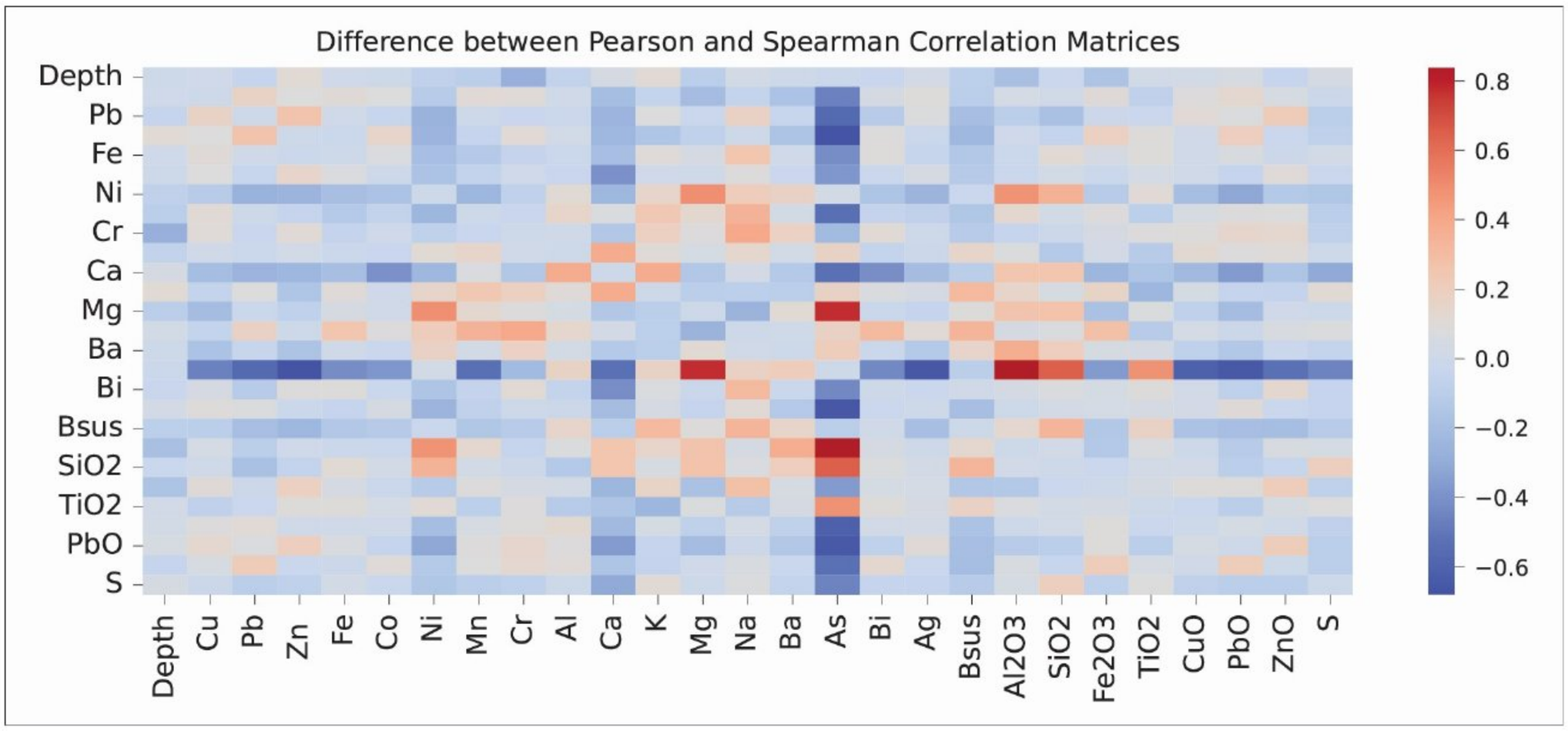

Figure 1. Selected differences between Pearson and Spearman correlation matrices.

An analysis of the cause behind the notable disparity in correlation coefficients for the same dataset revealed intriguing findings.Specifically, during the process of visualizing the data, a data point that deviated significantly from the rest of the data appeared, causing the data to be skewed. The PCC, which assumes a linear relationship between the data, failed to accurately estimate the true correlation coefficient when an outlier data point was present. In contrast, SCC, which does not make any assumptions about the distribution of the data or the type of relationship with the data, provided a more accurate estimation of the correlation coefficient.

Since this study was a preliminary investigation before further research in the field, it can be used as a reliable reference for evaluating the relationship between mineralogical, geophysical, and chemical data. Performing the concurrent computation of PCC, SCC and their discrepancy can provide a more comprehensive examination. The analysis can identify variable combinations that warrant further investigation and those that can be disregarded. If the PCC and SCC closely align, it indicates a stable correlation coefficient estimate. Otherwise, a more thorough examination is required.

CONCLUSION

Correlation coefficients are a powerful tool to asses type and strength of two or more variables relationship when utilized properly. In this brief communication the discrepancy between PCC and SCC was displayed and a general guideline on how to proceed was also given: If the PCC and SCC are closely aligned, it suggests a reliable estimate of the correlation between the variables. Alternatively, a more comprehensive analysis is necessary.

ACKNOLEDGEMENT

This research was supported by the Science Fund of the Republic of Serbia, 7522, – REASONING.

REFERENCES

1. Abramović, F., Cvetkov, V., Ilić, A., Životić, D. (2022). Korelacija magnetnog susceptibiliteta i sadržaja metala u flotacijskom jalovištu. XVIII Kongres geologa Srbije, Divčibare, 1-4 jun 2022.
2. Schober, P., Boer, C. and Schwarte, L.A., 2018. Correlation coefficients: appropriate use and interpretation. Anesthesia & Analgesia, 126(5), 1763-1768.