iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38121

Predict the amount of Cu using the four Ca, Al, P, S Elements by Multiple Linear Regression Method

Hamed Nazerian[1], Adel Shirazy[2], Aref Shirazi[3], Ardeshir Hezarkhani[4]
[1]University of Catania, [2, 3, 4]Amirkabir University of Technology

Abstract: The study area is located in Sarbisheh city in South Khorasan province, Iran. Copper estimation was performed by multivariate linear regression method to facilitate the use of previous analyses to predict this element in other areas, reduce costs and also reduce the number of samples. For this purpose, by obtaining a basic formula from estimating the amount of Cu with one of the promising points samples, the amount of copper in other parts of the exploration area was investigated. Several analyses were taken from the exploratory area after calculations to validate the regression. The regression results of new and old data were compared and estimation acceptable. These calculations were performed by SPSS software, according to the four elements Ca, Al, P, S, the results obtained and the relationship presented has acceptable validity.
Keywords: Multivariate linear regression, Cu estimation, SPSS, Iran.

I. INTRODUCTION

In statistical modeling, regression analysis [1-8] is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome' or 'response' variable) and one or more independent variables (often called 'predictors', 'covariates', 'explanatory variables' or 'features') [8-12]. The most common form of regression analysis is linear regression, in which one finds the line (or a more complex linear combination) that most closely fits the data according to a specific mathematical criterion. Regression analysis is primarily used for two conceptually distinct purposes. First, regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Second, in some situations regression analysis can be used to infer causal relationships between the independent and dependent variables. Importantly, regressions by themselves only reveal relationships between a dependent variable and a collection of independent variables in a fixed dataset. To use regressions for prediction or to infer causal relationships, respectively, a researcher must carefully justify why existing relationships have predictive power for a new context or why a relationship between two variables has a causal interpretation. The latter is especially important when researchers hope to estimate causal relationships using observational data. Also regression has various usage in earth science [13-23].

Important methods of regression analysis are as follows [24]:

1) Simple linear regression
2) Multiple linear regression
3) Fuzzy regression
4) logistic regression

Due to the high cost of exploration and also the need to save time, studies were conducted to reduce costs and time. In these decisions, a regression model was proposed to estimate the amount of copper elsewhere to use the analyses of the available area, as well as faster and cheaper future analyses.

There are many softwares that have the ability to calculate regression and the most famous of them are :

a) Excel (which is the simplest software)
b) SPSS
c) + S
d) SAS
e) R

In this case, due to the multivariate regression, SPSS software was used [25, 26].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

II. CU ESTIMATE USING SPSS SOFTWARE

In most engineering problems, there is more than one variable to describe the changes of a dependent variable. In these problems, linear or nonlinear multivariate regression method is used. The general model of the linear relationship is as follows:

$$y = \beta_1 + \beta_2 x_2 + ... + \beta_n x_n + e \qquad (1)$$

where in:

β is regression coefficients that are indeterminate, x are independent variables, y is a dependent variable, and e is a random error. Often, there is a difference between the actual points and the processed regression line. The difference between the observed points of the regression line (predicted value) is called the residual value. Lower residual values indicate a better prognosis. A good correlation coefficient is used to validate a good multivariate linear relationship [27-29].

In this study, SPSS 16 software was used to determine the unknown regression coefficients. In determining the amount of copper (dependent variable) by multivariate regression, the variables x1, x2, ... and xn as independent and 4 elements (Ca, Al, P, S) are dependent. In this study, used from 70 available data related to the results of analysis in the study area and 5 data from other areas. the statistical summary of the parameters is given in Table 1.

Table 1. Minimum and maximum values of the used data

Elements	Min (ppm)	Max (ppm)
Ca	36800	707000
Al	32300	103000
P	55	36600
S	50	3860

To examine the correlation of elements on the amount of copper in Table 2, the correlation between the 4 elements is given. According to the table 2, the four elements Ca, Al, P, S have the highest correlation on the amount of copper.

Table 2. Correlation matrix between variables

	Cu	Al	Ca	P	S
Cu	1				
Al	-0/74	1			
Ca	-0/47	-0/548	1		
P	-0/152	-0/484	0/963	1	
S	0/946	-0/470	0/932	0/877	1

The results of the statistical study are presented in Table 3. As shown in this table, Fisher's test was used to control the significance of the model and t-test was used to control the significance each of the independent variables.

Table 3. Statistical results of multivariate linear regression model for estimating copper content

Model	Unstandardized Coefficients		Standardized Coefficients	t	F
	B	Std. Error	Beta		
(Constant)	13829/43	2703/864		5/115	
P	3/781	0/487	0/583	7/769	
S	6/551	1/741	0/278	3/762	95/98
Ca	-0/117	0/016	-0/241	-7/248	
Al	-0/111	0/031	-0/127	-3/553	

a. Dependent Variable: Cu

According to Table 3 of the coefficients, the multivariate linear equation for estimating the amount of copper is as follows:

$$Cu = 13829/43 + 3/78P + 6/55S + (-0/117Ca) + (-0/11Al) \qquad (2)$$

By placing the data in the above equation, it is observed that as it was clear from the correlation coefficient, the estimation obtained from this equation is not much different from the real values and their difference is small and acceptable.

Table 4 summarizes the validity of the model. Figure 1 shows the actual values versus the estimated values. Considering the correlation coefficient of about 96% and the average error of about 4% for estimating the amount of copper, the proposed model indicates acceptable validity for prediction.

Table 4: Statistical Summary results of regression model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0/966[a]	0/933	0/929	0/025

a. Predictors: (Constant), Al, Ca, P, S

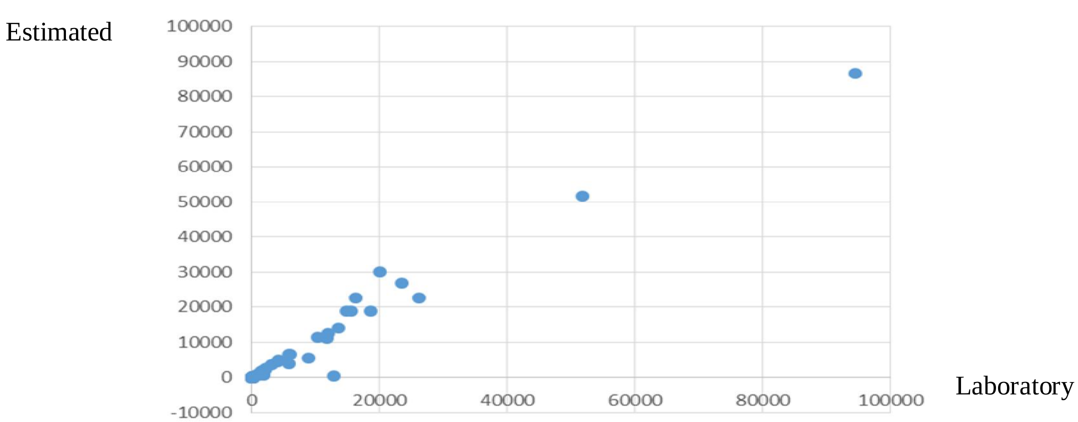

Figure 1: Estimated values versus laboratory values

Figure 2 also shows the comparison of estimated data with 5 other samples taken from other parts of the area.

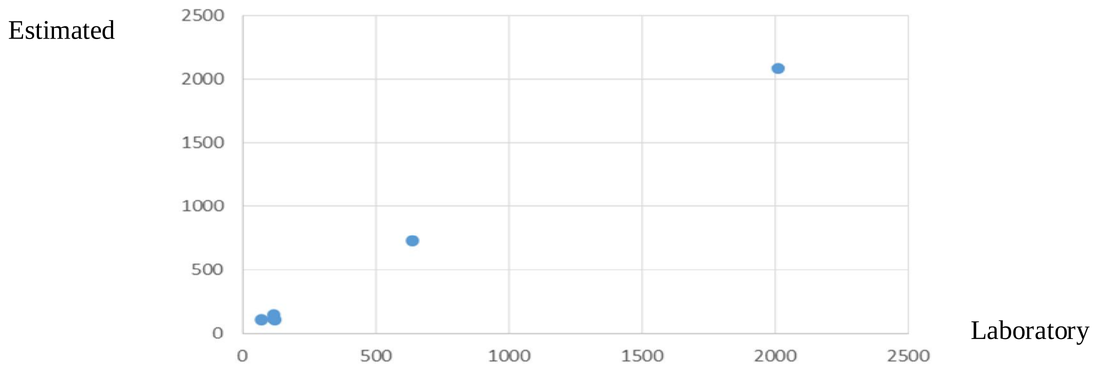

Figure 2: Estimated values versus laboratory values from other new points

III.CONCLUSIONS

The study area is in Sarbisheh city located in South Khorasan province, Iran. A total of 70 mineral samples from the study area were analyzed. With SPSS software, correlation test samples were taken from these 70 samples. The amount of copper was estimated using other S4 elements. Using Fisher test and t-test, the analysis was used for significance. Due to the significance of the results, the regression function was confirmed. Five other samples were taken from the area and analyzed. By comparing the estimated and the value of new samples, the error percentage was calculated to be 4%. by 96% accuracy of the regression, the results can be considered valid.

REFERENCES

[1] [Adel, S., Z. Mansour, and H. Ardeshir, Geochemical behavior investigation based on k-means and artificial neural network prediction for titanium and zinc, Kivi region, Iran. Известия Томского политехнического университета. Инжиниринг георесурсов, 2021. **332**(3): p. 113-125.
[2] Alahgholi, S., A. Shirazy, and A. Shirazi, Geostatistical studies and anomalous elements detection, Bardaskan Area, Iran. Open Journal of Geology, 2018. **8**(7): p. 697-710.
[3] Shirazi, A., et al., Exploration Geochemistry Data-Application for Cu Anomaly Separation Based On Classical and Modern Statistical Methods in South Khorasan, Iran. International Journal of Science and Engineering Applications, 2018. **7**: p. 39-44.
[4] Shirazi, A., et al., Geostatistics studies and geochemical modeling based on core data, sheytoor iron deposit, Iran. Journal of Geological Resource and Engineering, 2018. **6**: p. 124-133.
[5] Shirazy, A., et al., Geochemical and geostatistical studies for estimating gold grade in tarq prospect area by k-means clustering method. Open Journal of Geology, 2019. **9**(6): p. 306-326.
[6] Shirazy, A., M. Ziaii, and A. Hezarkhani, Geochemical Behavior Investigation Based on K-means and Artificial Neural Network Prediction for Copper, in Kivi region, Ardabil province, IRAN. Iranian Journal of Mining Engineering, 2020. **14**(45): p. 96-112.
[7] Shirazy, A., et al., Geostatistical and remote sensing studies to identify high metallogenic potential regions in the Kivi area of Iran. Minerals, 2020. **10**(10): p. 869.
[8] Jafari, S., et al. Expert pruning based on genetic algorithm in regression problems. in Asian Conference on Intelligent Information and Database Systems. 2012. Springer.
[9] Chatterjee, S. and A.S. Hadi, Regression analysis by example. Vol. 607. 2006: John Wiley & Sons.
[10] Draper, N.R. and H. Smith, Applied regression analysis. Vol. 326. 1998: John Wiley & Sons.
[11] Freund, R.J., W.J. Wilson, and P. Sa, Regression analysis. 2006: Elsevier.
[12] Seber, G.A. and A.J. Lee, Linear regression analysis. Vol. 329. 2012: John Wiley & Sons.
[13] Khayer, K., et al., Determination of Archie's Tortuosity Factor from Stoneley Waves in Carbonate Reservoirs. International Journal of Science and Engineering Applications (IJSEA), 2021. **10**: p. 107-110.
[14] Shirazi, A., et al., Remote sensing studies for mapping of iron oxide regions, South of Kerman, Iran. International Journal of Science and Engineering Applications, 2018. **7**(4): p. 45-51.
[15] Shirazi, A. and A. Shirazy, Introducing Geotourism Attractions in Toroud Village, Semnan Province, Iran. International Journal of Science and Engineering Applications, 2020. **9**(6): p. 79-86.
[16] Shirazi, A., A. Shirazy, and J. Karami, Remote sensing to identify copper alterations and promising regions, Sarbishe, South Khorasan, Iran. International Journal of Geology and Earth Sciences, 2018. **4**(2): p. 36-52.
[17] Shirazi, A., et al., Introducing a software for innovative neuro-fuzzy clustering method named NFCMR. Global Journal of Computer Sciences: theory and research, 2018. **8**(2): p. 62-69.
[18] Shirazy, A., et al., Exploratory Remote Sensing Studies to Determine the Mineralization Zones around the Zarshuran Gold Mine. International Journal of Science and Engineering Applications, 2018. **7**(9): p. 274-279.

[19] Shirazy, A., A. Shirazi, and A. Hezarkhani, Predicting gold grade in Tarq 1: 100000 geochemical map using the behavior of gold, Arsenic and Antimony by K-means method. Journal of Mineral Resources Engineering, 2018. **2**(4): p. 11-23.
[20] Shirazy, A., A. Shirazi, and H. Nazerian, Application of Remote Sensing in Earth Sciences–A Review. International Journal of Science and Engineering Applications, 2021. **10**: p. 45-51.
[21] Shirazy, A., et al., Geophysical study: Estimation of deposit depth using gravimetric data and Euler method (Jalalabad iron mine, kerman province of IRAN). Open Journal of Geology, 2021. **11**(8): p. 340-355.
[22] Ширази, А., et al., ИССЛЕДОВАНИЕ ГЕОХИМИЧЕСКОГО ПОВЕДЕНИЯ ТИТАНА И ЦИНКА НА ОСНОВЕ МЕТОДА К-СРЕДНИХ И ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ НОВЫХ ПЛОЩАДЕЙ, РЕГИОН КИВИ, ИРАН. Izvestiya Tomskogo Politekhnicheskogo Universiteta Inziniring Georesursov, 2021. **332**(3): p. 113-125.
[23] Shirazy, A., et al., Investigation of Geochemical Sections in Exploratory Boreholes of Mesgaran Copper Deposit in Iran. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2021. **9**(8): p. 2364-2368.
[24] Allen, M.P., Understanding regression analysis. 2004: Springer Science & Business Media.
[25] Frey, F., SPSS (software). The International Encyclopedia of Communication Research Methods, 2017: p. 1-2.
[26] Shirazy, A., Behavior of Gold, Arsenic, and Antimony Elements: K-Means Method, Matlab, and SPSS-Software. 2020: GRIN Verlag.
[27] Doodran, R.J., et al., Minimalization of Ash from Iranian Gilsonite by Froth Flotation. Journal of Minerals and Materials Characterization and Engineering, 2020. **9**(1): p. 1-13.
[28] Khakmardan, S., et al., Evaluation of Chromite Recovery from Shaking Table Tailings by Magnetic Separation Method. Open Journal of Geology, 2020. **10**(12): p. 1153-1163.
[29] Khakmardan, S., et al., Copper oxide ore leaching ability and cementation behavior, mesgaran deposit in Iran. Open Journal of Geology, 2018. **8**(09): p. 841.