

بررسی عوامل موثر بر کانی‌سازی مس با کمک رگرسیون تک متغیره، چند متغیره و رگرسیون لجستیک

ساناز حقیقی^۱، تکتام پورشیرازی^۲

^۱ دانشجوی کارشناسی ارشد مهندسی صنایع، دانشگاه فردوسی مشهد؛ sanazhaghighi336@yahoo.com

^۲ دانشجوی کارشناسی ارشد مهندسی صنایع، دانشگاه فردوسی مشهد؛ t.pourshirazi@gmail.com

چکیده

در مدل‌های آماری، تحلیل رگرسیون یک فرآیند آماری برای تخمین روابط بین متغیرها می‌باشد. تحلیل رگرسیون خصوصاً کمک می‌کند دریابیم چگونه مقدار متغیر وابسته با تغییر هر کدام از متغیرهای مستقل و ثابت بودن سایر متغیرها تغییر می‌کند. رگرسیون هم‌چنین به صورت گسترده‌ای برای پیش‌بینی استفاده می‌شود. برخی از مهم‌ترین شیوه‌های تحلیل‌های رگرسیونی از این قرار می‌باشد: رگرسیون خطی ساده، رگرسیون خطی چندگانه، رگرسیون فازی و رگرسیون لجستیک. در این مطالعه عوامل موثر بر کانی‌سازی مس با کمک تحلیل‌های رگرسیون که توسط کدنویسی در نرم‌افزار R به دست آمده، بررسی شده است. تمرکز این تحقیق بر روی ده متغیر خاص می‌باشد. وجود معادن متعدد مس در ایران و صادرات آن به سایر کشورها باعث شده تا ایران تأمین‌کننده حدود ۳ درصد از ذخایر مس دنیا باشد. با شناسایی عوامل موثر بر روی کانی‌سازی مس می‌توان با سرعت و دقت بیشتری فرآیند جستجو برای معادن مس را پیش برد.

کلمات کلیدی

تحلیل‌های رگرسیون، نرم‌افزار R، کانی‌سازی مس

Investigating the factors affecting copper mineralization with the help of single-variable regression, multivariate and logistic regression

Sanaz Haghighi, Toktam Pourshirazi

Masters student, Ferdowsi University Of Mashhad, Iran, sanazhaghighi336@yahoo.com

Masters student, Ferdowsi University Of Mashhad, Iran, t.pourshirazi@gmail.com

ABSTRACT

In statistical models, regression analysis is a statistical process for estimating the relationships between variables. Regression analysis specifically helps to find out how the value of the dependent variable varies by varying each of the independent variables and the constant of other variables. It is also widely used for prediction. Some of the most important methods of regression analysis are: Simple linear regression, multiple linear regression, fuzzy regression, and logistic regression. In this study, the effective factors on copper mineralization with the help of regression analysis, which was obtained by coding in software R, was investigated. The focus of this research is on ten specific variables. The existence of copper mines in Iran and its export to other countries has led Iran to supply about 3 percent of the world's copper reserves. By identifying the effective factors on copper mineralization, it is possible to quickly and accurately advance the search for copper mines.

KEYWORDS

Regression Analysis, R Software, copper mineralization

1- مقدمه

مدل سازی کنند. در این مقاله از مدل های رگرسیون چندگانه، فاکتوریل، چندجمله ای و رگرسیون سطح پاسخ استفاده گردیده است. مقاله بعدی با عنوان "پیش بینی پتانسیل معدن کاری با استفاده از داده های ژئوشیمیایی و مدل رگرسیونی خطی چندگانه در کانی سازی کربن کوه پنجه (ایران)" [2] است که در سال 2014 منتشر گردیده است. در این مطالعه، ترکیبی از تجزیه و تحلیل حفاری هسته با داده های ژئوشیمیایی سطح، به منظور تعیین موقعیت واقعی معدن مس انجام گرفته است. برای دستیابی به این هدف، از روش های رگرسیون چندمتغیره استفاده گردیده است. نتایج به دست آمده از این تحقیق برای مقایسه توزیع مس و مولیبدن استفاده گردید.

2- تحلیل واریانس رگرسیون تک متغیره

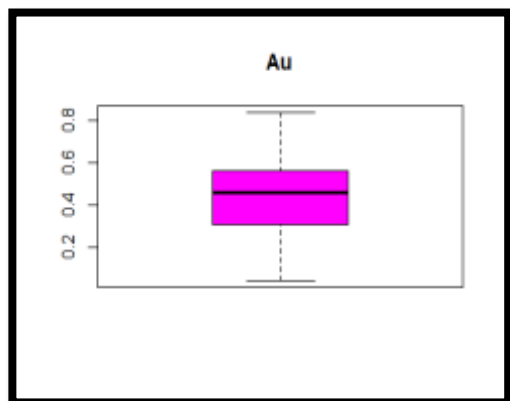
در ابتدا لازم است داده ها در نرم افزار تعریف شوند. نام هر کدام از متغیرها در نرم افزار به قرار جدول زیر می باشد:

جدول (1): معرفی متغیرها

نام متغیر در نرم افزار	نام عامل	نماد
Au	آنالیز طلا	x_1
Geology	نقشه زمین شناسی	x_2
Iron	بارزسازی اکسید آهن بر اساس تصاویر ماهواره ای	x_3
Argil	آلتراسیون آرژلیک بر اساس تصاویر ماهواره ای	x_4
Pro	آلتراسیون پروپلیتیک بر اساس تصاویر ماهواره ای	x_5
Factor.1	فاکتور مرتبط با کانی سازی بر اساس آنالیز فاکتوری داده - های ژئوشیمیایی	x_6
Mo	آنالیز مولیبدن	x_7
Total	نقشه میدان کل مغناطیسی	x_8
RTP	نقشه برگردان به قطب داده های مغناطیسی	x_9
Fault	گسل های محدوده	x_{10}
cu	کانی سازی مس	y

با توجه به اهمیت کانه مس و کاربرد وسیع آن در صنایع متفاوت و با توجه به محدود بودن ذخیره معادن فعال موجود، شناخت مناطق دارای پتانسیل مناسب بسیار ضروری به نظر می رسد. در همین راستا و با توجه به مشخص شدن و به اثبات رسیدن تأثیر عوامل ساختاری و پارامترهای دیگر بر روی کانی سازی مس، می توان در روند شناسایی معادن مس تغییرات مثبتی اعمال نمود. با کمک تحلیل های رگرسیون ثابت می شود که آیا متغیرها بر روی کانی سازی موثر می باشند یا خیر. عوامل ساختاری و پارامترهای موثر بر کانی سازی مس توسط تحقیقی که بر روی یک معدن مس در جنوب استان خراسان انجام گردید از این قرار می باشند: آنالیز طلا، نقشه زمین شناسی، بارزسازی اکسید آهن بر اساس تصاویر ماهواره ای، آلتراسیون آرژلیک بر اساس تصاویر ماهواره ای، آلتراسیون پروپلیتیک بر اساس تصاویر ماهواره ای، فاکتور مرتبط با کانی سازی بر اساس آنالیز فاکتوری داده های ژئوشیمیایی، آنالیز مولیبدن، نقشه میدان کل مغناطیسی، نقشه برگردان به قطب داده های مغناطیسی و عامل دهم گسل های محدوده می باشد.

1-1- مرور ادبیات: در رابطه با پیشینه تحقیق مورد نظر جست و جوی های زیادی صورت گرفت. برای بررسی عوامل موثر بر کانی سازی مس پیشینه تحقیق غنی ای وجود نداشت. در ادامه سه مورد از مواردی که بیشترین ارتباط را با موضوع مورد نظر داشتند به اختصار ذکر شده است. اولین مقاله با عنوان "پیش بینی ماکزیمم سرعت ذرات با استفاده از شبکه های عصبی مصنوعی و تجزیه و تحلیل رگرسیون چند متغیره - معدن مس سرچشمه، کرمان، ایران" [2] می باشد. در این تحقیق که در سال 2013 منتشر شده است، شاخص PPV (ماکزیمم سرعت ذرات) در کرمان با استفاده از سه روش: شبکه عصبی مصنوعی، تحلیل رگرسیون چندمتغیره و ارتباطات تجربی، پیش بینی و مقایسه شده است. این مقاله دارای 5 متغیر مستقل می باشد که شامل 30 مجموعه داده تشکیل شده از 24 مجموعه تست و 6 مجموعه آزمون است. در نهایت در نتیجه مقایسه ای که با استفاده از نتایج تحلیلات سه روش به دست آمد، مشخص شد که شبکه های عصبی به علت ماهیت غیرخطی بودن، انعطاف پذیری بالا و خطای کم، دارای اعتبار بیشتری می باشد. مقاله دوم مربوطه به تحقیق مورد نظر، مقاله ای با عنوان "تجزیه و تحلیل رگرسیون چند متغیری از داده های ژئوشیمیایی به منظور مدل سازی کانه زایی زیرزمین: مطالعه موردی از رسوب طلای ساری، شمال غربی ایران" [2] است که در سال 2014 منتشر گردید. در این تحقیق، رگرسیون چندمتغیره بر روی داده های ژئوشیمیایی به دست آمده از ذخایر طلای یکی از معادن طلای ساری به کار گرفته شد تا با استفاده از آن بتوانند کانی سازی زیرزمینی را برای حفاری پیش تر



شکل (2): نمودار جعبه‌ای آنالیز طلا

همبستگی بین دو متغیر عبارت است از میزان تأثیری که این دو متغیر بر روی یکدیگر دارند.

جدول (3): همبستگی خطی بین آنالیز طلا و کانی‌سازی مس

	cu	Au
cu	1.0000000	0.6590652
Au	0.6590652	1.0000000

میزان همبستگی بین این دو متغیر نشان‌دهنده رابطه‌ای نسبتاً قوی بین این دو می‌باشد. پس از بررسی داده‌های متغیر مستقل به صورت مجزا نوبت آن است که به فرمولی مناسب برای بیان رابطه بین این دو متغیر دست یابیم. نمای کلی مدل از این قرار است:

$$cu = \beta_0 + \beta_1 Au$$

با تعریف مدل در نرم‌افزار و به کمک دستور summary mod پارامترهای مدل و سایر اطلاعات جدول زیر به دست می‌آید.

جدول (4): اطلاعات دستور Summary

b_0	0.14037
b_1	0.84496
α	0.05
F -statistic	766.4
$F(0.05, 1, 998)$	0.003934118
P-value	2.2e-16

جدول (5): حدود مقادیر باقی‌مانده‌ها

Min	-0.21234
1Q	-0.13232
Median	-0.06376
3Q	0.12472
Max	0.42349

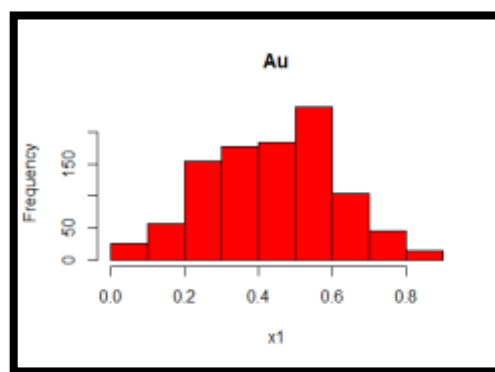
با توجه به این که مقدار p-value برای متغیر Au بسیار کمتر از مقدار احتمال خطای نوع اول می‌باشد، متغیر مذکور به شدت بر روی کانی‌سازی مس موثر می‌باشد. از منظر دیگر مشاهده می‌شود که مقدار

متغیر پاسخ کانی‌سازی مس در نظر گرفته می‌شود، زیرا لازم است تا تأثیر سایر متغیرها بر روی آن بررسی و تحلیل شود. کار با رگرسیون تک متغیره آغاز می‌شود. همان‌طور که از نامش پیداست در این نوع رگرسیون تنها تأثیر یک عامل بر روی متغیر پاسخ بررسی می‌شود. در ادامه به تحلیل تأثیر متغیر آنالیز طلا روی کانی‌سازی پرداخته خواهد شد.

جدول (2): تعدادی از داده‌ها

Cu	Au
0.5675750	0.5887195
0.8455778	0.4628122
0.7301324	0.5467955
0.7646488	0.4433047
0.4778446	0.1500518
0.8102824	0.6162025

نمودار هیستوگرام با توجه به داده‌ها رسم می‌شود و با توجه به شکل این نمودار می‌توان توزیع مناسب داده‌ها را تخمین زد.



شکل (1): نمودار هیستوگرام آنالیز طلا

از شکل نمودار مشخص است که داده‌ها به توزیع نرمال نزدیک می‌باشند.

با کمک نمودار جعبه‌ای می‌توان به مرکزیت داده‌ها پی برد و هم-چنین داده‌ها را از لحاظ میزان پراکندگی بررسی نمود. از مشاهده نمودار جعبه‌ای مشخص است، همان‌طور که از نمودار هیستوگرام نتیجه شد، داده‌ها به توزیع نرمال نزدیک بوده و داده پرت وجود ندارد.

برای آزمون فرض زیر نتایج از این قرار می‌باشد:

$$\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$$

جدول (6): نتایج آزمون فرض β_1

t.test	t.alpha
907.0162	-1.962344

با توجه به بزرگتر بودن مقدار آماره آزمون فرض صفر رد می‌شود.

2-2- تحلیل‌های β_0 :

با احتمال 0.95٪ مقدار β_0 در داخل این محدوده قرار دارد:

$$\beta_0 \in (0.1120836, 0.1686484)$$

برای آزمون فرض زیر نتایج از این قرار می‌باشد:

$$\begin{cases} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{cases}$$

جدول (7): نتایج آزمون فرض β_0

t.test	t.alpha
150.6744	-1.962344

به همان دلیل فرض صفر در مورد پارامتر نخست نیز رد می‌شود.

2-3- پیش‌بینی یک مشاهده جدید: در بخش مقدمه اشاره مختصری به کاربردهای رگرسیون شد. یکی از مهم‌ترین کاربردهای آن پیش‌بینی یک مشاهده جدید برای متغیر پاسخ می‌باشد. در اینجا فرض شده است مقدار متغیر x_1 برابر با جمع مقدار بیشینه و کمینه آن قرار داده می‌شود و سپس مقدار کانی‌سازی مس برابر می‌شود با:

$$0.8853824$$

2-4- ضریب تعیین: $R^2 = 0.4343$ این مقدار هرچه به عدد یک نزدیک باشد نشان از کفایت بیشتر مدل دارد. در این مورد هرچند مدل برازش داده شده تا حدودی مناسب است اما قطعاً بهترین نمی‌باشد.

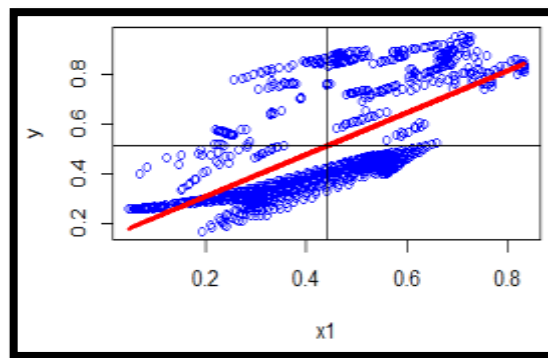
2-5- ضریب همبستگی: این مقدار همواره عددی است بین منفی و مثبت یک. این مقدار برای مسأله مورد نظر ما $r = 0.6590652$ به دست آمده است که نشان از یک رابطه خطی مثبت دارد. همچنین رابطه زیر برقرار می‌باشد:

$$r^2 = 0.6590652^2 = 0.4343669 = R^2$$

2-6- آزمون نرمال بودن: برای این آزمون به‌طور شهودی به نمودار

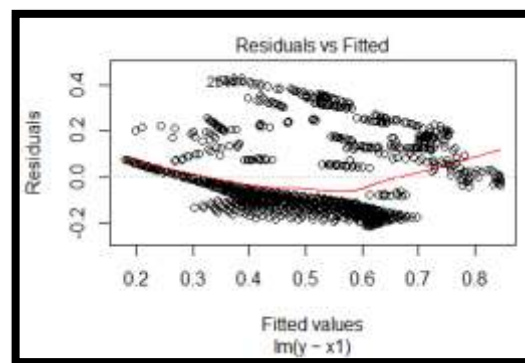
زیر دقت کنید:

نوزیع F مدل بیشتر از $F(0.05, 1, 998)$ می‌باشد. این بدین معناست که این مدل کاملاً مناسب می‌باشد.



شکل (3): نمودار برازش داده شده بین متغیر پاسخ و مستقل

در این نمودار خط حاصل از تقاطع دو بعد $(\bar{x}, \bar{y})$ می‌باشد. ملاحظه می‌شود که خط رگرسیون از میانگین x و میانگین y عبور می‌کند. از نمودار باقی‌مانده‌ها در برابر مقادیر برازش داده شده می‌توان دریافت که آیا رابطه‌ای بین متغیرها برقرار می‌باشد و همچنین آیا واریانس خطا ثابت است یا خیر.



شکل (4): نمودار باقیمانده‌ها در برابر مقادیر برازش داده شده

همانطور که مشاهده می‌شود فرم پراکنش باقیمانده‌ها الگوی خاصی را نشان نمی‌دهد بنابراین یک رابطه خطی بین متغیر پاسخ و مستقل برقرار می‌باشد و گویای ثابت بودن واریانس خطا نیز است.

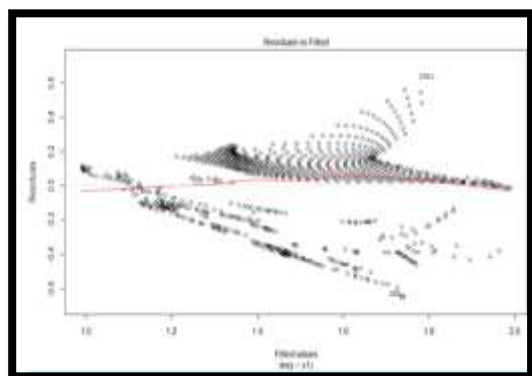
2-1- تحلیل‌های β_1 :

با احتمال 0.95٪ مقدار β_1 در داخل این محدوده قرار دارد:

$$\beta_1 \in (0.7850680, 0.9048568)$$

از روش ترسیمی و همچنین برخی از آزمون های آماری استفاده شده است.

اولین نمودار مورد بررسی نمودار باقیمانده ها در مقابل مقادیر برازش داده شده می باشد که این نمودار به صورت زیر است:



شکل (7): نمودار باقیمانده ها در مقابل مقادیر برازش داده شده

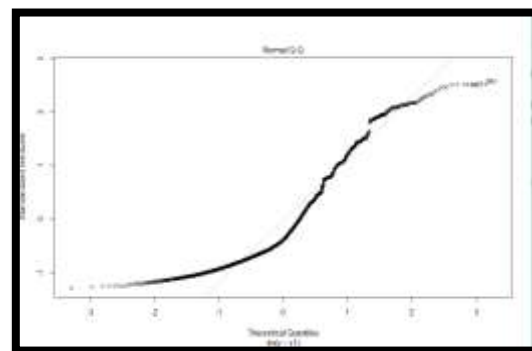
با توجه به این نمودار فرض مناسب بودن مدل خطی تا حدودی برابر است اما شاید فرض برابری واریانس با توجه به شکل تا حدی رد شود برای بررسی بیشتر از آزمون لیون اصلاح شده برای پی بردن به این نکته استفاده شده است. نتایج حاصل از این آزمون در جدول زیر آمده است:

جدول (8): نتایج آزمون لیون اصلاح شده

ناحیه پذیرش	$[-2.33, 2.33]$
خطای نوع اول	0.01
مقدار آماره	2.007
P-value	0.045

با توجه به مقدار آماره 2.007 و بازه پذیرش فرض برابری واریانس ها پذیرفته می شود. همچنین با توجه به مقدار p-value این آزمون که مقدار 0.045 دارد، برابری واریانس ها در سطح 0.01 پذیرفته می شود.

برای بررسی نرمال بودن باقیمانده ها نیز از نمودار احتمال نرمال بودن باقیمانده ها رسم می شود. با توجه به این نمودار مشاهده می شود که بعد از ایجاد تبدیل وضعیت نرمال بودن بهتر شده و به صورت تقریبی می توان فرض نرمال بودن باقیمانده ها را پذیرفت.

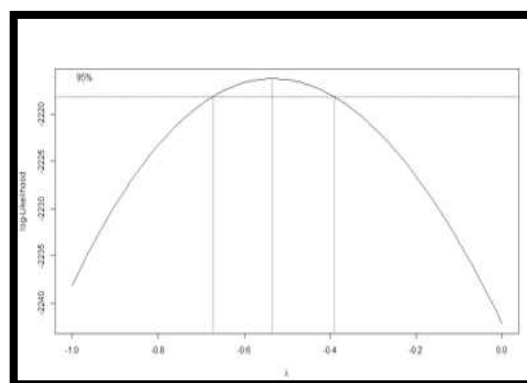


شکل (5): آزمون نرمال بودن

همان گونه که ملاحظه می شود با توجه به فاصله داشتن باقیمانده ها از خط، فرض نرمال بودن رد می شود، همچنین برای بررسی نرمال بودن باقیمانده ها از 2 آزمون شاپیرو و کلموگروف-اسمیرنف استفاده شده است که خروجی های این 2 آزمون نیز نشان دهنده نرمال نبودن باقیمانده ها می باشد.

2-6-1-آزمون شاپیرو: مقدار P-value این آزمون برابر شده با $2.2e-16$ ، که مقداری بسیار کم می باشد. بنابراین با توجه به خروجی این آزمون نیز فرض نرمال بودن باقی مانده ها رد می شود.
2-6-2-آزمون کلموگروف-اسمیرنف: بر اساس این آزمون مقدار P-value برابر شده با $2.2e-16$ که مقداری بسیار اندک می باشد. در نتیجه با توجه به این آزمون نیز فرض نرمال بودن باقی مانده ها رد می شود.

2-7-تصحیح مدل: همان گونه که در قسمت قبل بررسی شد فرض نرمال بودن باقیمانده ها رد می شود، برای حل این مشکل از روش تبدیل باکس-کاکس استفاده شده است.

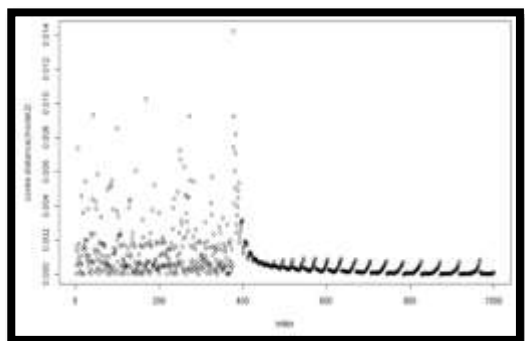


شکل (6): تبدیل مناسب باکس-کاکس

با توجه به نتیجه نرم افزار مقدار پیشنهادی برای تبدیل روی متغیر پاسخ تقریباً عدد -0.5 عدد مناسبی می باشد. پس از انجام این تبدیل مجدداً نیاز است که کفایت مدل بررسی شود. برای بررسی کفایت مدل

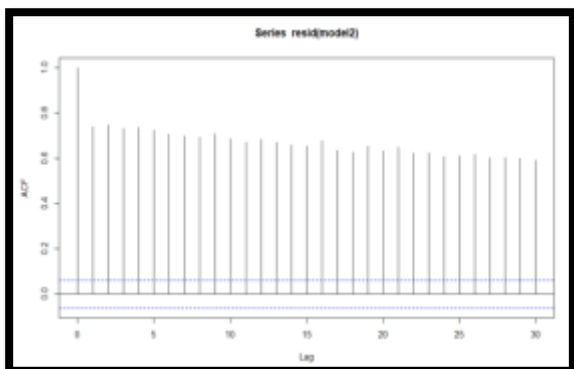
شکل (10): نمودار مقادیر کلاه

3-7-2- بررسی وجود داده تاثیرگذار: برای این منظور از فاصله کوک استفاده می‌شود. با توجه به مقدار بحرانی 0.69 و این که هیچ کدام از این مقادیر بزرگتر از این عدد نمی‌باشند پس داده تاثیرگذار وجود ندارد.



شکل (11): نمودار فاصله کوک

4-7-2- بررسی وجود خودهمبستگی: همان طور که مشخص است با توجه به نمودار AFC مقادیر هم‌چنان خارج از حدود قرار دارند، بنابراین مشکل خودهمبستگی بین باقی مانده‌ها هم‌چنان برقرار می‌باشد.



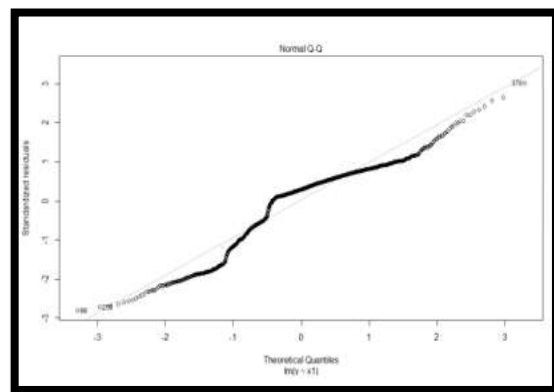
شکل (12): نمودار AFC

8-2- پیش‌بینی معکوس: مدل رگرسیونی Y نسبت به X برای پیش‌بینی سطحی از X در ازای مشاهده نویی از Y قرار می‌گیرد. به طور مثال اگر مقدار Y را برابر با جمع مقدار کمینه و بیشینه آن قرار دهیم در نتیجه مقدار X برار خواهد شد با:

$$\hat{x}_{k(new)} = 1.168019$$

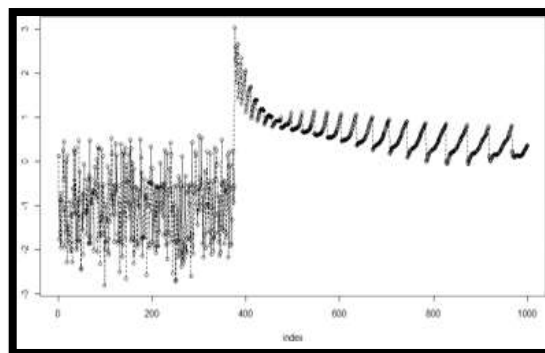
9-2- آزمون همبستگی: هدف این آزمون استنباط در خصوص صفر بودن همبستگی جمعیت می‌باشد. آزمون فرض ما به فرم زیر در نظر گرفته شده است:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$



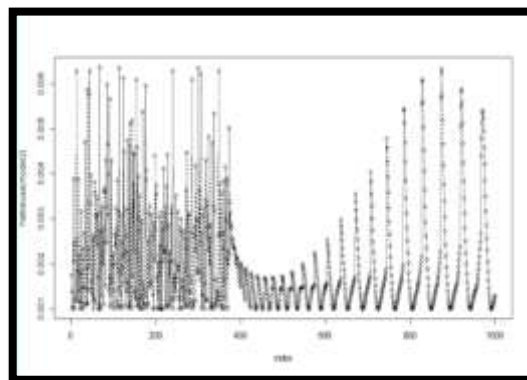
شکل (8): آزمون نرمال بودن

1-7-2- بررسی وجود داده پرت: برای بررسی وجود داده پرت از نمودار باقیمانده استاندارد حذف شده استفاده می‌شود. مقدار حد بحرانی برای این نمودار بازه $[-4.07, 4.07]$ می‌باشد و با توجه به این که همه داده‌ها در این بازه قرار دارند پس می‌توان به این نتیجه رسید که داده پرتی وجود ندارد.



شکل (9): نمودار باقیمانده استاندارد حذف شده

2-7-2- بررسی وجود داده اهرمی: برای شناسایی نقاط اهرمی از مقادیر ماتریس کلاه و یا نمودار مقادیر کلاه استفاده می‌شود. با توجه به مقدار بحرانی 0.04 تعداد زیادی از نقاط به عنوان نقاط اهرمی شناخته می‌شوند.



جدول (9): نتایج آزمون همبستگی

t.test	t.alpha
150.6744	-1.962344

همانطور که مشاهده می شود مقدار $150.6744 > -1.962344$ شده است. بنابراین فرض H_0 رد می شود. به عبارت دیگر یک رابطه خطی بین متغیرها برقرار می باشد.

3- تحلیل واریانس رگرسیون چندمتغیره

مسائل زیادی وجود دارند که برای پیش بینی بهتر رفتار متغیر وابسته، لازم است بیش از یک متغیر مستقل در نظر بگیریم. پس از بررسی متغیر آنالیز طلا بر روی کانی سازی مس به صورت مدل رگرسیون تک متغیره، نوبت آن است تا نتیجه تاثیر هم زمان تمام ده متغیر به صورت هم زمان بر روی متغیر پاسخ بررسی شود.

در ابتدا با رسم نمودار هیستوگرام و هم چنین نمودار جعبه ای برای هر سری از داده ها می توان به یک سری اطلاعات اولیه دست یافت. توزیع متغیرهای x_3 تا x_6 به تابع توزیع نرمال نزدیک می باشد.

با دقت در نمودار جعبه ای متغیر $geology$ مشاهده می شود که داده ها دارای چولگی می باشند و هم چنین داده پرتی در داده ها دیده نمی شود. داده های مربوط به متغیر $iron$ دارای پراکندگی نرمال و متقارن می باشند و هیچ گونه چولگی در داده ها وجود ندارد و هم چنین با دقت در نمودار جعبه ای متغیر $argil$ مشاهده می شود که داده ها دارای پراکندگی نرمال و تقریباً متقارن می باشند و مقداری چولگی در پراکندگی داده ها دیده می شود، دو داده پرت در پراکندگی وجود دارد. داده های متغیر pro دارای پراکندگی نرمال و متقارن هستند و هیچ گونه چولگی در داده ها وجود ندارد، هم چنین داده پرتی در مشاهده نمی شود. با دقت در نمودار جعبه ای متغیر $factor.1$ مشاهده می شود که داده ها دارای پراکندگی نرمال می باشند. مقداری چولگی و تعدادی داده پرت در پراکندگی داده ها دیده می شود. در مورد متغیر Mo مقداری چولگی در پراکندگی داده ها و هم چنین تعدادی داده پرت دیده می شود. با دقت در نمودار جعبه ای متغیر $Total$ مشاهده می شود که چولگی زیاد و تعدادی داده ی بسیار پرت در پراکندگی داده ها دیده می شود. در مورد داده های متغیر RTP چولگی زیادی در پراکندگی داده ها دیده می شود. هم چنین داده پرت وجود ندارد. و در انتها با دقت در نمودار جعبه ای متغیر $fault$ مشاهده می شود که چولگی زیادی در پراکندگی داده ها وجود دارد و هم چنین تعدادی داده ی بسیار پرت مشاهده می شود.

3-1- برازش مدل: مدل با وجود تمام متغیرهای بیان شده وهم-

چنین اثرات درجه 2

$$Cu = \beta_0 + \beta_1 Au + \beta_2 geology + \beta_3 iron + \beta_4 argil + \beta_5 pro + \beta_6 factor.1 + \beta_7 Mo + \beta_8 Total + \beta_9 RTP + \beta_{10} fault$$

جدول (10): پارامترهای مدل چندگانه

پارامتر	برآورد
β_0	0.18
β_1	0.033
β_2	0.35
β_3	-0.0034
β_4	-0.262
β_5	-0.325
β_6	0.831
β_7	0.079
β_8	-0.210
β_9	0.391
β_{10}	-0.082

مقدار p.value متغیرهای $geology$ و $argil$ و pro و $factor.1$ و $Total$ و RTP و $fault$ برازش از مقدار 0.05 بسیار کوچکتر شده است. بنابراین این متغیرها در مدل به شدت تاثیر گذار هستند.

جدول (11): مقادیر P-value متغیرها

متغیر	P-value
x_1	0.336
x_2	2e-16
x_3	0.9
x_4	3.44e-12
x_5	2e-16
x_6	2e-16
x_7	0.102
x_8	1.37e-06
x_9	1.68e-10
x_{10}	2.65e-05

جدول (12): ضرایب تعیین و همبستگی

ضریب تعیین چندگانه	$R^2 = 0.88$
ضریب تعیین چندگانه تعدیل یافته	$R_a^2 = 0.8788$
ضریب همبستگی چندگانه	$\sqrt{R^2} = r = 0.9380832$

جدول (13): جدول تحلیل واریانس

SST	47.90752
SSR	-51.02395

$$Cu = \beta_0 + \beta_1 Au + \beta_2 geology + \beta_3 iron + \beta_4 argil + \beta_5 pro + \beta_6 factor.1 + \beta_7 + \beta_8 Total + \beta_9 RTP + \beta_{10} fault + \beta_{11} Au * geology + \dots$$

تعدادی از ضرایب در جدول زیر آمده است. متغیرهای سوم، ششم، هشتم و دهم در سطح 0.05 معنی‌دار شناخته شده‌اند. همچنین اثر متقابل بین متغیر مستقل اول و سوم نیز معنی‌دار شناخته شده است. مقدار ضریب تعیین چندگانه برابر است با 0.9722 و مقدار ضریب تعیین چندگانه تعدیل یافته برابر 0.9706 می‌باشد. چون مقدار p -value: $< 2.2e-16$ می‌باشد. بنابراین مدل برای داده‌ها مناسب می‌باشد و می‌توان فرض صفر بودن همزمان همه متغیرها را رد نمود.

جدول (16): مدل رگرسیون با اثرات متقابل

متغیر	پارامتر	p-value
x_1	1.16143	2.29e-05
x_1	1.3735	4.05e-05
x_3	-0.0071	0.9732
x_4	0.4263	0.0402
x_5	0.6740	4.40e-07
x_6	0.2749	0.4519
x_7	2.8626	5.57e-09
x_8	-1.3136	0.06576
x_9	5.7379	2e-16
x_{10}	-0.0499	0.8095
$x_2 x_1$	-1.0972	0.00072
$x_3 x_1$	-0.2742	0.1601
$x_4 x_1$	-1.7736	1.69e-11
$x_5 x_1$	0.5191	0.01835
$x_6 x_1$	-1.3806	2e-16
$x_7 x_1$	0.7682	0.06412
$x_8 x_1$	1.9798	0.00014
$x_9 x_1$	-0.7600	0.2737

98.93147	SSE
----------	-----

جدول (14): فواصل اطمینان پارامترها

عرض از مبدا	0.11293636	0.25079321
x_1	-0.03481561	0.10195920
x_2	0.27897709	0.42610299
x_3	-0.05702871	0.05016014
x_4	-0.33497471	-0.18903881
x_5	-0.38184280	-0.26948152
x_6	0.73188385	0.93137278
x_7	-0.01581155	0.17536405
x_8	-0.29601658	-0.12571663
x_9	0.27243589	0.51035060
x_{10}	-0.12040811	-0.04398917

ضریب جزیی برای Cu و Au زمانی که $geology$ در مدل موجود

است برابر است با

$$R_{r,1.2}^2 = 0.5780487$$

و ضریب جزیی برای Cu و $geology$ زمانی که Au در مدل موجود

است برابر است با:

$$R_{r2.1}^2 = 0.6974958$$

2-3- بررسی همبستگی خطی چندگانه: برای بررسی وجود و یا عدم

وجود همبستگی خطی چندگانه از معیار VIF که همان عامل تورم

واریانس می‌باشد، استفاده شده است. خروجی این روش به صورت زیر

می‌باشد. چون ماکسیمم این مقادیر بزرگ‌تر از 10 می‌باشد پس می‌-

توان گفت مشکل همبستگی خطی چندگانه وجود دارد.

جدول (15): معیار VIF

متغیر	معیار
x_1	6.089136
x_2	1.861886
x_3	1.825672
x_4	2.953794
x_5	3.964987
x_6	9.481360
x_7	6.970171
x_8	7.943593
x_9	12.384531
x_{10}	1.926590

3-3- مدل رگرسیون با اثرات متقابل: مدل مورد نظر ما با اثرات

متقابل به فرم زیر می‌باشد:

β_7	0.08188162
β_8	-5.38511168
β_9	51.04025
β_{10}	2.57318

مشاهده می‌شود که متغیرهای Au و geology و factor.1 و RTP در مدل به شدت تأثیرگذار می‌باشند و متغیر pro نیز نسبت به 4 متغیر دیگر تأثیر کمتری در مدل می‌گذارد. باقی متغیرها در این مدل تأثیر گذار نیستند.

5- جمع‌بندی و نتیجه‌گیری

عوامل ساختاری و پارامترهای موثر بر کانی‌سازی مس توسط تحقیقی که بر روی یک معدن مس در جنوب استان خراسان انجام گردید مشخص گردید. سپس با تحلیل‌های رگرسیون اطلاعات متفاوتی درباره انواع روابط بین این متغیرها به دست آمد. در مدل تک متغیره مشخص شد که باقی‌مانده‌ها از توزیع نرمال پیروی نمی‌کنند، از این رو با کمک تبدیل باکس-کاکس این مشکل تا حد زیادی برطرف گردید. در این مدل داده پرت و تأثیرگذار وجود نداشت، اما تعداد زیادی داده اهرمی مشاهده می‌شد.

در تحلیل واریانس چند متغیره مشخص شد که از بین تمام متغیرها، geology و argil و pro و factor.1 و Total و RTP و fault در مدل بسیار تأثیرگذار می‌باشند. با بررسی بیشتر مشخص گردید که در این مدل نیز مشکل همبستگی خطی چندگانه وجود دارد. در برازش مدل با اثرات متقابل نیز متغیرهای شماره سه، شش، هشت و ده معنی‌دار می‌باشند و همچنین اثر متقابل بین متغیر مستقل اول و سوم نیز معنی‌دار شناخته شده است. نتیجه این تحقیق قطعا در نهایت باعث تسریع روند کشف معادن مس در ایران خواهد شد.

6-مراجع

- [2] H. Bakhshandeh Amnieh1*, A. Mohammadi1, M. Mozdianfard2, 2013, Predicting peak particle velocity by artificial neural networks and multivariate regression analysis - Sarcheshmeh copper mine, Kerman, Iran
- [3] Hamid Graniana, Seyed Hassan Tabatabaei, Hooshang H. Asadi, Emmanuel John M. Carranza, 2014, Multivariate regression analysis of lithogeochemical data to model subsurface mineralization: a case study from the Sari Gunay epithermal gold deposit, NW Iran

$x_{10}x_1$	-0.8363	0.00040
x_3x_2	-0.5265	0.02300
x_4x_2	0.5585	0.04090
x_5x_2	-0.5511	0.07680
x_6x_2	0.1620	0.6786

3-4- روش جستجوی اتماتیک به روش‌های رو به جلو و روش رو به عقب: خروجی این روش نمایش متغیرهای مناسب برای برازش بهترین مدل با استفاده از روش forward می‌باشد. مشاهده می‌شود که تمام اثرات درجه 1 معنی‌دار شناخته شده‌اند و همچنین بین برخی از متغیرها نیز اثر متقابل معنی‌داری وجود دارد. در روش رو به عقب نیز تمامی اثرات درجه 1 معنی‌دار شناخته می‌شوند.

4- تحلیل واریانس رگرسیون لجستیک

رگرسیون لجستیک می‌تواند به عنوان یک مورد خاص از مدل خطی عمومی در نظر گرفته شود. این مدل بر اساس فرض‌های کاملاً متفاوتی (درباره رابطه متغیرهای وابسته و مستقل) از رگرسیون خطی است. تفاوت مهم این دو مدل در دو ویژگی رگرسیون لجستیک نمایان می‌شود. اول این که توزیع شرطی در لجستیک یک توزیع برنولی می‌باشد، زیرا متغیر وابسته دودویی است دوم این که مقادیر پیش‌بینی احتمالاتی است، همچنین محدود بین بازه صفر و یک و به کمک تابع توزیع لجستیک بدست می‌آید. [1]

در این مسأله فرض شده است که مقدار Cu به صورت باینری باشد، به گونه ای که داده‌های بزرگتر از 0.5 با کد 1 و کوچکتر از 0.5 با کد 0 در نظر گرفته شود. بنابراین متغیر پاسخ یک متغیر باینری می‌باشد. رگرسیون لجستیک به فرم زیر برازش داده می‌شود:

$$cu \sim x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}$$

,family = binomial

جدول (16): برآورد پارامترهای مدل رگرسیون

پارامتر	برآورد
β_0	-79.02406
β_1	-81.94879
β_2	40.76118503
β_3	-9.73534489
β_4	4.40165785
β_5	11.60209287
β_6	173.51340523

Seyed Hasan Tabatabaei · Parisa Roshani Rodsari ·
Ahmad Reza Mokhtari, 2014, Predicting Potential
Mineralization Using Surface Geochemical Data and
Multinle Linear Regression Model in the Kuh Pani

[4]

[5]

[www. wikipedia.org](http://www.wikipedia.org)