

ماهیت ترکیبی داده‌های لیتوژئوشیمیایی و تاثیر آن بر تحلیل رگرسیونی عیار طلا: مطالعه موردی اندیس طلای قره‌چر، شمال غرب ایران

حسین ملایمت^۱، فرهاد محمدتراب^{۲*}، امیرحسین کوهساری^۳

۱- دانشجوی دکتری، دانشکده مهندسی معدن و متالورژی، دانشگاه یزد

۲- دانشیار، دانشکده مهندسی معدن و متالورژی، دانشگاه یزد

۳- دانشیار، دانشکده مهندسی معدن و متالورژی، دانشگاه یزد

(دریافت: اسفند ۱۳۹۴، پذیرش: دی ۱۳۹۷)

چکیده

تحلیل رگرسیونی به عنوان بخش مهمی از اکتشاف ذخایر طلا با داده‌های ژئوشیمیایی با ماهیت ترکیبی سر و کار دارد. در داده‌های ترکیبی، متغیرهای عیاری بخشی از کل بوده و اطلاعات مفید در آنها در قالب نسبت متغیرها وجود دارد. محدود بودن داده‌ها در بازه صفر تا ۱۰۰ درصد باعث ایجاد همبستگی تحمیلی و کوواریانس دروغین شده و این موضوع، تحلیل‌های آماری مرسوم را با اشتباه و عدم قطعیت همراه خواهد کرد. این موضوع برای داده‌های ژئوشیمیایی که مجموع آنها به ۱۰۰ درصد نمی‌رسد نیز ثابت شده است. در این پژوهش تحلیل رگرسیونی عیار طلا در یک رخداد معدنی طلا در شمال غربی ایران با دو رویکرد سنتی (رگرسیون غیرخطی چند متغیره داده‌های خام) و ترکیبی (رگرسیون غیرخطی چند متغیره بالانس‌های به دست آمده از تبدیل نسبت لگاریتمی ایزومتريک (ilr)) مورد توجه قرار گرفته است. هدف از این کار مقایسه توانایی دو رویکرد در برآورد دقیق عیار طلا است. ضریب همبستگی مقادیر واقعی و تخمینی برای رویکرد سنتی ۰/۷۴ و برای رویکرد ترکیبی ۰/۸۴ و نرخ طبقه‌بندی صحیح (CCR) به عنوان مثال در عیار ۴۰ppb به ترتیب ۰/۶۸۸ و ۰/۸۷۵ است. به علاوه، تابع توزیع مقادیر تخمینی با رویکرد ترکیبی بیش از رویکرد سنتی به تابع توزیع داده‌های واقعی شبیه است. همچنین، رویکرد سنتی در یک نمونه، عیار طلا را کم‌تر از صفر تخمین زده است. این تناقض آشکار نشانه‌ای از نامناسب بودن روش‌های سنتی ریاضی برای کار با داده‌هایی با سیستم بسته است. به علاوه، رویکرد ترکیبی در برآورد چارک‌های اول و سوم که مقادیر حیاتی‌تری در جدایش داده‌های آنومال هستند، تخمین‌های دقیق‌تری ارائه نموده است. مجموع این موارد دلالت بر این دارند که به منظور یک تحلیل آماری دقیق و قابل اعتماد، باید ماهیت ترکیبی داده‌های ژئوشیمیایی مورد توجه واقع شده و داده‌ها با روش نسبت لگاریتمی مناسب تحلیل شوند.

کلید واژه‌ها

داده ترکیبی، تبدیل نسبت لگاریتمی ایزومتريک ilr، رگرسیون چند متغیره، فاصله ماهالانوبیس، طلا، قره‌چر

۱- مقدمه

سبب وجود این ویژگی‌ها، نباید از روش‌های آماری مانند تحلیل فاکتوری، تحلیل تمایز و تحلیل مؤلفه اصلی که بر مبنای ماتریس کوواریانس و یا ماتریس ضرایب همبستگی هستند، برای تحلیل داده‌های ترکیبی خام استفاده نمود [۷-۹]. با این وجود هنوز هم در پروژه‌های اکتشافی توجه عمده‌ای به این مشکلات نمی‌شود. شایان ذکر است که در مواردی، به کارگیری روش‌های آماری در تحلیل داده‌های ترکیبی خام ظاهراً نتایج قابل تفسیر و معقولی به دست می‌دهد. هر چند این نتایج به ظاهر معقول، با عدم قطعیت غیرقابل مدل‌شدن همراه هستند.

اولین بار چایس ثابت کرد که همبستگی موجود بین اجزاء یک دسته داده ترکیبی، غیرواقعی است. وی با مطالعات دقیق‌تر نشان داد که این همبستگی ناشی از رفتار مشابه عناصر در فرآیندهای فیزیکوشیمیایی تمرکز دهنده در طبیعت نبوده و صرفاً به خاطر ماهیت ترکیبی داده‌ها به وجود می‌آید [۹]. پس از چایس پژوهشگران دیگری نیز منشا همبستگی اجزاء داده‌های ترکیبی را مورد توجه قرار دادند [۱۰، ۱۱]. در نهایت مشخص شد که نه تنها همبستگی‌های ژئوشیمیایی در داده‌های ترکیبی حقیقی نیستند، بلکه سایر روش‌های آماری برای تحلیل این داده‌ها با اشکالات بنیادی روبرو هستند [۱۲]. آیچیسون علت بروز این مشکلات را کشف و این‌گونه بیان نمود که داده‌های ترکیبی حاوی اطلاعاتی در قالب نسبت‌های عددی هستند و نه مقادیر مطلق.

در این پژوهش ابتدا روش‌های نوین تحلیل داده‌های ترکیبی برای عبور از مشکلات طرح شده در بالا معرفی می‌شوند. در ادامه، با در نظر گرفتن ویژگی‌های داده‌های لیتوزئوشیمیایی و با توجه به هدف قرار دادن تحلیل رگرسیونی عیار طلا، رویکرد تحلیل داده‌های ترکیبی مناسب اتخاذ می‌شود. به منظور بررسی تاثیر این رویکرد بر دقت و قابلیت اطمینان مدل رگرسیونی، نتایج مدل رگرسیون ترکیبی با مدل رگرسیون سنتی مقایسه می‌شود. با در نظر گرفتن معیارهای اعتبارسنجی مختلف، در نهایت مشخص می‌شود که در نظر گرفتن ماهیت ترکیبی داده‌های لیتوزئوشیمیایی (و به طور عام ژئوشیمیایی) در تحلیل‌های آماری چند متغیره به چه میزان می‌تواند نتایج این تحلیل‌ها را بهبود بخشد.

در ژئوشیمی و اکتشافات معدنی، داده‌های ترکیبی به صورت بخشی از کل تعریف و معمولاً به صورت بردارهایی از درصدها، غلظت‌ها و به طور کلی، اجزاء بیان می‌شوند. از آنجایی که این اجزاء به صورت اعداد حقیقی نمایش داده می‌شوند، این باور اشتباه وجود دارد که می‌توان آنها را داده‌های چندمتغیره در نظر گرفت و به تجزیه و تحلیل آماری آنها اقدام نمود [۱]. به کارگیری رابطه‌های ریاضی حاکم بر داده‌های حقیقی در مورد داده‌های ترکیبی می‌تواند منجر به ایجاد تفسیر اشتباه و تناقض شود. این موضوع از یک قرن پیش شناخته شده است اما عموماً فراموش شده و طی سال‌های گذشته در بیشتر مطالعات ژئوشیمیایی نادیده گرفته شده است [۲].

مثال آشنا برای یک سیستم عددی بسته، نتایج آنالیز شیمیایی مواد معدنی است که متغیرهای آن مستقل از یکدیگر نبوده و به صورت درصد یا قسمت در میلیون (ppm) بیان می‌شوند. در گذشته، تنها مجموعه داده‌های با حاصل جمع ثابت به عنوان داده‌های ترکیبی شناخته می‌شدند اما در حال حاضر مفهوم داده ترکیبی دارای تعریف وسیع‌تری است و داده‌هایی که دارای حاصل جمع ثابت نباشند را نیز شامل می‌شود [۳].

داده‌های ترکیبی دارای خواصی هستند که باعث می‌شوند استفاده از روش‌های آماری استاندارد برای تحلیل آنها به صورت خام با عدم قطعیت و خطا و اشتباه همراه باشد [۴]. روش‌های آماری استاندارد برای پردازش داده‌هایی که در بازه $-\infty$ تا $+\infty$ تغییر می‌نمایند، طراحی شده‌اند [۵، ۶]. داده‌های ترکیبی همیشه مثبت بوده و هنگامی که به شکل بسته هستند فقط در بازه صفر تا ۱۰۰ یا هر ثابت دیگری تغییر می‌کنند. مشکل آنجاست که پس از تحلیل‌های چندمتغیره مشخص نمی‌شود چه مقدار از عدم قطعیت مدل نهایی مربوط به تحمیلی بودن همبستگی و کوواریانس‌ها است.

متغیرهای داده‌های ترکیبی بسته نمی‌توانند به صورت مستقل تغییر کنند و تغییرات آنها وابسته به یکدیگر است. در این صورت اگر یک متغیر افزایش یابد، سایر متغیرها به اجبار کاهش می‌یابند. این موضوع قویاً می‌تواند نتایج تحلیل‌های آماری و تخمین‌ها را دچار اریب‌شدگی کند. به

۲- تحلیل داده‌های ترکیبی

آن وجود دارد. در این تبدیل هر متغیر بر ریشه حاصل ضرب متغیرهای دیگر تقسیم می‌شود، سپس از مقادیر به دست آمده لگاریتم گرفته و در آخر در ضربی که وابسته به تعداد متغیرهاست ضرب می‌شود. پس از انجام این تبدیل بر روی داده‌ها، تعدادی بالانس به دست می‌آید و کلیه محاسبات و تحلیل‌های آماری بر روی این بالانس‌ها انجام و در نهایت نتایج به فضای بسته اولیه تبدیل معکوس می‌شوند. نکته مهم این است که پس از انجام این تبدیل، ارتباط خطی بین متغیرهای اولیه و بالانس‌ها از بین می‌رود [۱۸].

با توجه به مجموع نکات اشاره شده در بالا، تبدیل ilr نسبت به دو تبدیل دیگر گزینه مناسبتری برای تحلیل رگرسیونی چندمتغیره عیار طلا تشخیص داده شد.

۳- کانسارهای طلای اپی‌ترمال و پهنه برشی

کانسارهای طلا را می‌توان بر اساس نوع محصول به دو گروه شامل کانسارهایی با محصول اصلی طلا و کانسارهایی با محصول فرعی طلا تقسیم کرد. هر یک از این دو گروه شامل بیش از ده تیپ کانسار است [۱۹]. در این میان، کانسارهای اپی‌ترمال کانسارهایی با محصول اصلی طلا هستند که امروزه به عنوان یک ذخیره معدنی جایگاه قابل توجهی پیدا کرده‌اند [۲۰]. عامل پیدایش این کانسارهای طلا، فعالیت‌های گرمایی اپی‌ترمال عنوان شده است. کانسارهای اپی‌ترمال محصول فعالیت سیستم‌های هیدروترمال بزرگی هستند که انرژی لازم برای فعالیت خود را از توده‌های ماگمای تزریق شده در بخش‌های بالایی پوسته زمین تامین می‌کنند.

در کانسارهای طلای پهنه برشی پراکنده در سنگ‌های آتشفشانی، طلا در زون آلتره به صورت ذرات خالص، الکتروم، همراه با پیریت، آرسنوپیریت و سینابر یافت می‌شود. ساختار غالب این نوع کانسارها رگه و رگچه‌ای، برشی، دانه پراکنده و بندرت استوک‌ورک است [۲۱، ۲۲]. رفتار ژئوشیمیایی متقابل عناصر در نمونه‌های اخذ شده از محیط سنگی یک کانسار طلای پهنه برشی مورد بحث این مقاله است.

به صورت اولیه، آرسنیک، آنتیموان و جیوه به دلیل رابطه زایشی با طلا معرف‌های خوبی هستند و ممکن است به عنوان ردیاب استفاده شوند. فلزات پایه (مس، سرب و

به منظور باز کردن داده‌های ژئوشیمیایی با سیستم بسته و استخراج اطلاعات مفید در قالب نسبت متغیرها، سه نسبت لگاریتمی توسط ریاضی‌دانان پیشنهاد شده است. تبدیل نسبت لگاریتمی افزوده ($additive\ logratio$) (alr) در ۱۹۸۶ توسط آیچیسون پیشنهاد شد. در این تبدیل از بین متغیرهای موجود، یک متغیر انتخاب و سایر متغیرها بر آن تقسیم شده، سپس از آنها لگاریتم گرفته می‌شود. بدین ترتیب سیستم از حالت بسته خارج می‌گردد و ارتباط نسبتی بین متغیرها بارز می‌شود. با این حال، تبدیل alr با مشکلاتی همراه است از جمله اینکه به کارگیری این تبدیل برای یک دسته داده واحد با جواب یکتا همراه نیست. چرا که نتایج به انتخاب متغیر تقسیم‌کننده وابسته هستند و این متغیر نیز بر اساس تجربه و نظر شخصی تعیین می‌گردد. تا کنون هیچ نظریه و روند استاندارد شده‌ای برای انتخاب این متغیر معرفی نشده است [۱۳]. به علاوه در سال‌های اخیر ثابت شده است که تبدیل alr نمی‌تواند همبستگی تحمیلی را به طور کامل از داده‌های بسته حذف نماید [۱۴، ۱۵].

پس از چندی، تبدیل نسبت لگاریتمی مرکزی ($centered\ logratio$) (clr) توسط آیچیسون پیشنهاد شد [۱۶]. در این تبدیل مقدار هر متغیر بر میانگین هندسی ردیفی (نمونه‌ای) که به آن تعلق دارد تقسیم شده و از عدد به دست آمده، لگاریتم طبیعی گرفته می‌شود. در مقایسه با تبدیل alr ، متغیری از مجموعه داده‌ها حذف نمی‌شود اما در مقابل، یکی از معایب تبدیل clr معکوس‌ناپذیر بودن ماتریس کوواریانس متغیرهاست، چون دترمینان این ماتریس برابر با صفر است. بنابراین بسیاری از تحلیل‌های آماری چندمتغیره اعم از تحلیل رگرسیونی که هدف این پژوهش است را نمی‌توان به راحتی بر روی داده‌های تبدیل یافته با clr انجام داد.

در ادامه و در ۲۰۰۳ تبدیل نسبت لگاریتمی ایزومتریک ($isometric\ logratio$) (ilr) توسط اگزکیو و همکاران مطرح شد [۱۷]. امکان محاسبه معکوس ماتریس کوواریانس در تبدیل ilr وجود دارد. علاوه بر این، از لحاظ خواص هندسی ترجیح داده می‌شود. در مقابل، تبدیل ilr نسبت به دو تبدیل قبلی پیچیده‌تر است و قواعد متفاوتی برای انجام

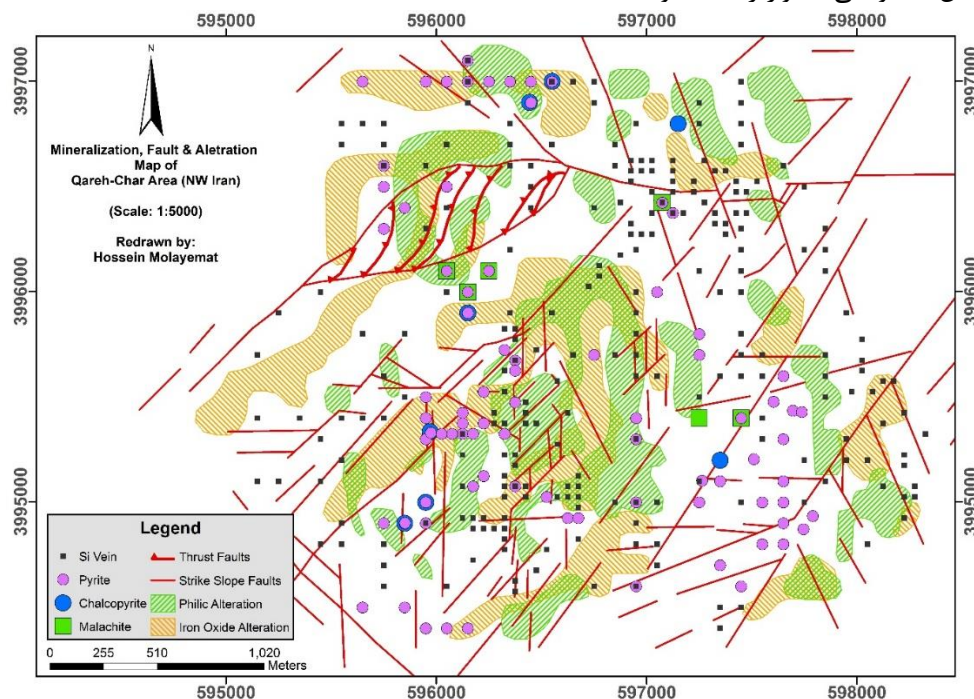
شیست‌های کلریت و اپیدوت‌دار و گرانیست است. در درون شیست‌های کلریت- اپیدوت‌دار پدیده و روند دگرگونی در یک زون شدیداً برشی صورت گرفته است، به طوری که در مطالعه میکروسکوپی نمونه‌های سنگی این واحد دگرگون شده به خوبی نشانگر همزمانی پدیده دگرگونی (افزایش گرادیان حرارتی) در یک پهنه برشی فعال در این گستره است [۲۴].

دگرسانی غالب در محدوده مورد مطالعه از نوع فلیک و اکسیدهای آهن است، که بخش وسیعی از محدوده را شامل می‌شود (شکل ۱). علی‌رغم گسترش زیاد واحدهای گنیس و گرانیست، کانه‌های فلزی کالکوپیریت، پیریت و کم‌تر ملاکیت به صورت پراکنده در این واحدها وجود دارند که می‌تواند به دلیل تهی بودن ماگمای گرانیستی اولیه از آغشتگی و آلودگی به ماده معدنی فلزی باشد. شبکه نمونه- برداری به روش لب‌پری به صورت سلول‌های مربعی به ابعاد ۵۰ متر (۵۱۵ سلول) و ۱۰۰ متری (۶۰۶ سلول) طراحی شده است. این نمونه‌ها به روش جذب اتمی برای ۱۰ عنصر As, Ag, Au, Bi, Cu, Mo, Pb, Sb, Sn و Zn در آزمایشگاه زرکاوان البرز آماده‌سازی و آنالیز شده‌اند.

روی) نیز ممکن است برای ردیابی و اکتشاف برخی از انواع نهشته‌های طلا مفید باشند [۲۳]. در بسیاری از کانسارهای طلا، عناصری به صورت پاراژنز حضور دارند که طی فرآیندهایی مشابه طلا پراکنده شده و یا تمرکز می‌یابند. معمولاً تجزیه این عناصر راحت‌تر و کم هزینه‌تر بوده و با دقت بالایی همراه است. مدلسازی آماری ارتباط ژنتیک و تخمین غلظت طلا بر اساس عیار این عناصر می‌تواند تا حد زیادی از هزینه‌های اکتشافات طلا کاسته و به اجرای این طرح‌ها سرعت بخشد. یکی از راه‌های شناخت ارتباط این عناصر با طلا، تحلیل رگرسیونی چندمتغیره است.

۴- رخدادهای معدنی قره‌چر

محدوده مورد مطالعه در استان کردستان و بین شهرهای سقز و بانه واقع شده است. این محدوده به وسعت ۱۰/۷ کیلومترمربع در برگه ۱/۱۰۰۰۰۰ سقز و برگه ۱/۵۰۰۰۰ میرده قرار گرفته است و از نظر ساختاری در پهنه سنج سیرجان و در شمال کمربند ساختاری بیستون قرار دارد. واحدهای سنگی موجود در محدوده گرانیست- گنیس میلونیتی، کوارتز- مسکویت شیست،



شکل ۱: نقشه پراکندگی گسل‌ها، دگرسانی‌ها، کانه‌زائی و موقعیت نمونه‌های لیتوژئوشیمیایی در محدوده معدنی قره‌چر [۲۵]

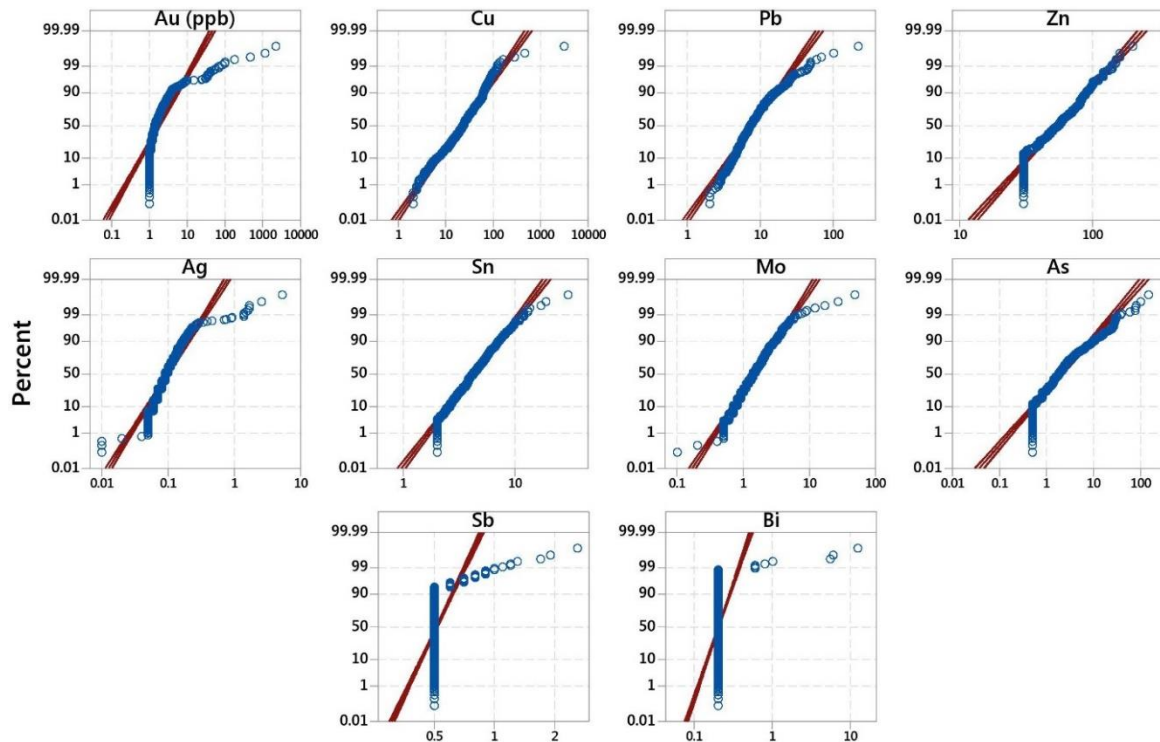
به منظور مطالعه تابع توزیع و پراکندگی داده‌ها، نمودارهای احتمال ۱۰ عنصر آنالیز شده رسم شدند (شکل ۲). مشخص است که تقریباً تمامی مقادیر عناصر بیسموت

۵- تحلیل رگرسیونی طلا

۵-۱- بررسی آماری اولیه

نسبت لگاریتمی ایزومتریکی^۱ جایگزین شدند (شکل ۳). این روش جایگزینی با حفظ ساختار کوواریانس میان متغیرها، روشی مناسب برای کار با داده‌های چند متغیره به حساب می‌آید [۲۶].

و آنتیموان سنسورد هستند. به همین دلیل این دو عنصر به طور کامل از داده‌ها کنار گذاشته شدند. به علاوه، تقریباً تمام عناصر به خصوص روی، طلا، آرسنیک و قلع دارای مقادیر سنسورد هستند. به منظور ادامه بررسی‌های آماری، مقادیر سنسورد با روش نوین و دقیق بیشینه امید ریاضی



شکل ۲: نمودار احتمال ۱۰ عنصر آنالیز شده بر حسب ppm (تقریباً تمامی مقادیر بیسموت و آنتیموان سنسورد هستند)

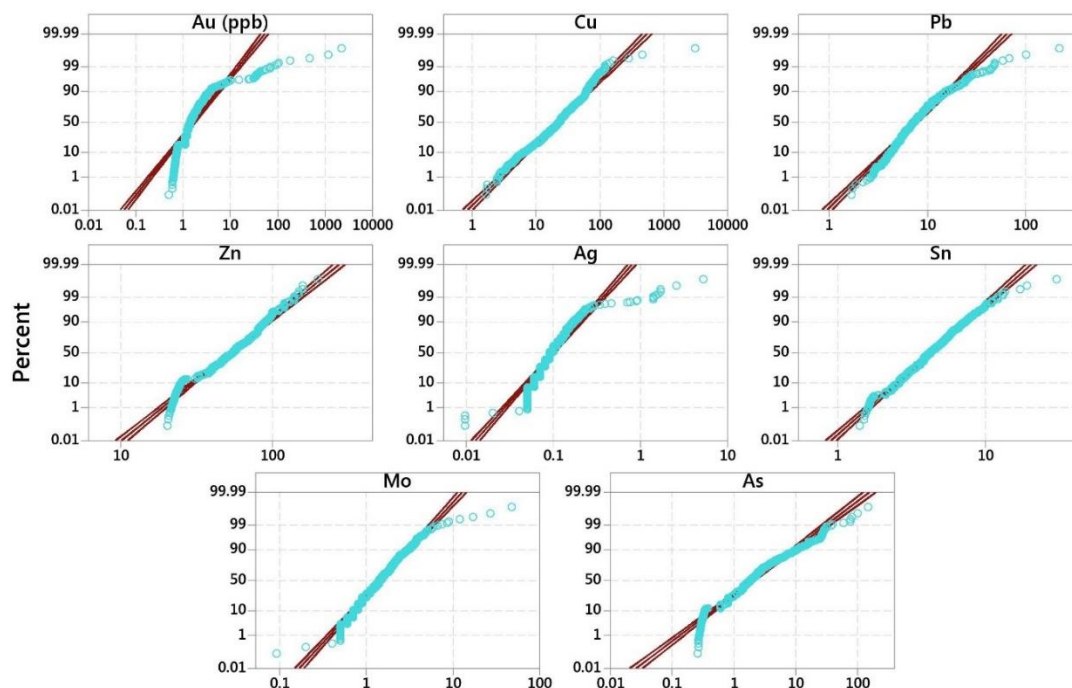
نماید. با اینکه این روش برای جدایش آنومالی در داده‌های چندمتغیره با اشکالاتی همراه است اما به دلیل سهولت اجرا و درک ریاضی آسان، کماکان در بسیاری از پروژه‌های اکتشافی کشور به کار گرفته می‌شود.

به منظور جامعیت پژوهش حاضر، تاثیر ماهیت ترکیبی داده‌ها بر تحلیل رگرسیونی نمونه‌های آنومال به دست آمده با هر دو روش بالا بررسی شده است. بر اساس روش فواصل ماهالانوبیس مقاوم (Robust) تعداد ۳۲ نمونه و بر اساس حد جدایش $\bar{x} + 2s$ (که برابر با مقدار ۱۵ ppb است) تعداد ۴۹ نمونه آنومال هستند. داده‌های مربوط به این نمونه‌ها در جدول‌های ۱ و ۲ آورده شده است. در مراحل بعدی، این نمونه‌ها مورد تجزیه و تحلیل رگرسیونی قرار گرفته‌اند.

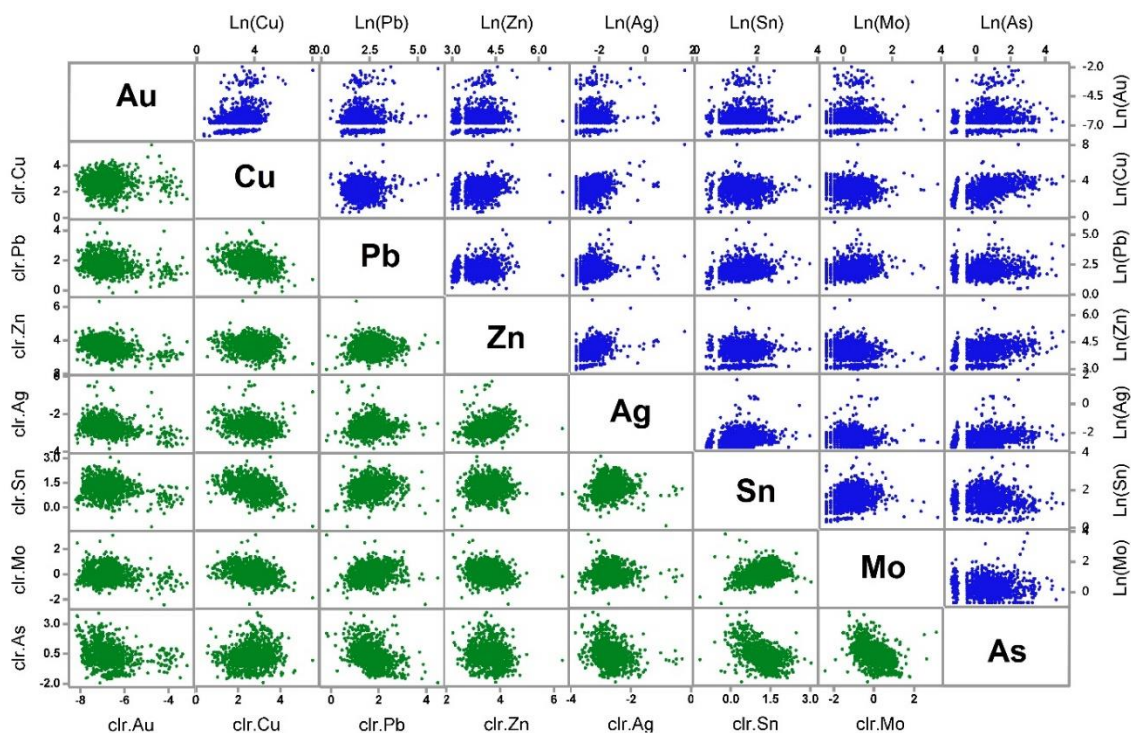
در ادامه این پژوهش از مقادیر غلظت عناصر Ag, As, Au, Cu, Mo, Pb, Sn و Zn برای ۱۱۲۱ نمونه لیتوژنوشیمیایی استفاده شده است. شکل ۴ بیانگر ارتباط دو به دوی عناصر، پیش و پس از باز کردن داده‌ها برای ۱۱۲۱ نمونه است.

۵-۲- جدایش آنومالی

در این پژوهش دو روش غیر ساختاری برای تشخیص نمونه‌های آنومال به کار گرفته شده است: روش چندمتغیره فواصل ماهالانوبیس و روش تک‌متغیره سنتی بر اساس مقدار $\bar{x} + 2s$. روش فواصل ماهالانوبیس بر اساس مقادیر تمام عناصر آنالیز شده در هر نمونه اقدام به جدایش نمونه‌های آنومال می‌نماید اما در انتها مشخص نمی‌شود که هر عنصر چه نقشی در آنومال شدن نمونه داشته است [۲۷]. در مقابل روش سنتی قرار دارد که تنها طلا را در نظر گرفته و با تعریف یک مقدار مرزی، نمونه‌های آنومال را معرفی می‌کند.



شکل ۳: نمودار احتمال ۸ عنصر پس از جایگزینی مقادیر سنسورد با روش ilr-EM (به جز طلا، مقادیر بر حسب ppm هستند)



شکل ۴: نمودار پراکنش عناصر نسبت به یکدیگر. مثلث بالا: پراکنش مقادیر لگاریتم طبیعی عناصر؛ مثلث پایین: پراکنش عناصر پس از باز شدن با تبدیل نسبت لگاریتم مرکزی (clr)

جدول ۱: غلظت عناصر مختلف در ۳۲ نمونه آنومال بر اساس روش فواصل ماهالانویس (بر حسب ppm)

Sample ID	Au (ppb)	Cu	Pb	Zn	Ag	Sn	Mo	As
425	140	23	36	96	0.11	6.6	1.4	6.5
606	119	100	420	590	0.52	5.6	0.7	2.9
265	103	3100	24	160	5.3	3.8	1.8	11.3
593	99	38	15	64	0.1	6	2.1	7.8
528	98	26	7.4	73	0.11	13	2.3	9.5
661	84.5	40	7.6	62	0.1	3.3	1.2	3.4

ادامه جدول ۱: غلظت عناصر مختلف در ۳۲ نمونه آنومال بر اساس روش فواصل ماهالانوبیس (بر حسب ppm)

Sample ID	Au (ppb)	Cu	Pb	Zn	Ag	Sn	Mo	As
592	70.5	32	7.8	68	0.12	6.7	2.6	7.8
633	68	6.6	5	80	0.06	5.3	0.8	1
853	68	55	9.6	73	0.19	3.4	1	8.4
480	66.5	14	12	57	0.15	8.1	1.4	1.5
610	62	110	5.4	66	0.12	3.8	0.7	3.7
918	56.5	280	9.6	70	0.76	4	1.6	2.7
490	56	15	14	190	0.09	11	2.3	1.4
589	54	44	5.8	73	0.19	5.7	1.1	2.3
637	53.5	36	18	50	0.09	3.3	3.2	5.4
590	52.5	33	9.6	58	0.12	4.8	2.1	2.7
951	52	44	6	40	0.08	4.4	0.8	2.1
1017	46.5	55	5.4	50	0.06	3.9	1.7	2.7
1078	45	16	7.3	36	0.08	2.8	1.4	0.33
1021	43	390	4.6	70	0.5	4.2	1.7	1.4
206	37.5	93	5.6	48	0.23	4.2	17	14.9
969	32	460	24	80	0.92	3.6	1.2	1.5
1040	6	92	5.4	63	0.19	4.3	1.4	39.5
291	5	29	13	60	0.11	42.5	1.7	2.5
910	4.6	32	6.5	96	1.5	6	4	2.7
1108	4.5	130	16	32	0.16	1.774	12	12.7
1041	3.8	60	5.4	50	0.14	3.8	2.5	59.6
947	2.7	36	22	89	1.4	7.6	3.8	6.4
858	2.7	16	34	60	0.17	19	5	3.3
955	2.5	64	68	150	0.12	4.8	2.1	77.2
1086	0.6106	60	10	200	0.1	3.6	1	26.4
1128	1.2	88	58	160	0.12	3.6	1.2	149

جدول ۲: غلظت عناصر مختلف در ۴۹ نمونه آنومال بر اساس مرز جدایش $\bar{x} + 2s$ (بر حسب ppm)

Sample ID	Au (ppb)	Cu	Pb	Zn	Ag	Sn	Mo	As
425	140	23	36	96	0.11	6.6	1.4	6.5
606	119	100	420	590	0.52	5.6	0.7	2.9
265	103	3100	24	160	5.3	3.8	1.8	11
593	99	38	15	64	0.1	6	2.1	7.8
528	98	26	7.4	73	0.11	13	2.3	9.5
661	84.5	40	7.6	62	0.1	3.3	1.2	3.4
592	70.5	32	7.8	68	0.12	6.7	2.6	7.8
633	68	6.6	5	80	0.06	5.3	0.8	1
853	68	55	9.6	73	0.19	3.4	1	8.4
480	66.5	14	12	57	0.15	8.1	1.4	1.5
610	62	110	5.4	66	0.12	3.8	0.7	3.7
918	56.5	280	9.6	70	0.76	4	1.6	2.7
490	56	15	14	190	0.09	11	2.3	1.4
589	54	44	5.8	73	0.19	5.7	1.1	2.3
637	53.5	36	18	50	0.09	3.3	3.2	5.4
590	52.5	33	9.6	58	0.12	4.8	2.1	2.7
951	52	44	6	40	0.08	4.4	0.8	2.1
571	47	66	27	76	0.15	8.2	0.8	8.5
1017	46.5	55	5.4	50	0.06	3.9	1.7	2.7
1078	45	16	7.3	36	0.08	2.8	1.4	0.3
868	44	30	6.2	60	0.05	2.4	2	2.5
1021	43	390	4.6	70	0.5	4.2	1.7	1.4
397	41.5	62	40	76	0.1	5	2.8	2.7
692	40.5	30	13	32	0.08	5.6	2.1	7
772	40	51	13	60	0.12	4.3	1.3	0.8
529	39	24	7.2	76	0.18	5.3	1.9	7
554	38.5	20	5.8	68	0.1	3.7	2.6	4.3
206	37.5	93	5.6	48	0.23	4.2	17	15
562	36.5	58	4.4	68	0.09	11	2.7	3.5
816	35.5	5.3	6.2	49	0.08	5.8	2.2	1.3
462	35	12	7.5	87	0.09	2.9	1.8	1

ادامه جدول ۲: غلظت عناصر مختلف در ۴۹ نمونه آنومال بر اساس مرز جدایش $\bar{x} + 2s$ (بر حسب ppm)

Sample ID	Au (ppb)	Cu	Pb	Zn	Ag	Sn	Mo	As
983	35	40	8.8	56	0.29	3.8	1.6	1.8
868	35	68	6	42	0.05	3.6	1.1	3
824	34.5	66	11	78	0.09	5	1.5	3.6
974	34.5	9.3	43	45	0.11	6.3	1.9	3.7
996	34.5	24	7	50	0.12	8.5	0.8	6.5
407	33	20	10	56	0.05	3.5	1.9	2
594	33	48	19	68	0.12	6.7	0.9	4.6
969	32	460	24	80	0.92	3.6	1.2	1.5
497	31.5	31	6.9	80	0.06	2.7	1.6	1.2
653	31.5	19	7.7	43	0.11	5.5	2.2	1.3
654	30.5	22	7.6	46	0.09	3.8	1.6	1.9
597	29	40	3.2	25	0.1	4.6	1.5	4.4
612	28.5	19	6.8	40	0.07	4.2	0.9	2.9
639	27.5	48	7.7	60	0.17	4.1	3.4	11
885	25.5	41	7.1	55	0.34	7.8	1.3	2.4
621	24	8	3.8	21	0.06	1.62	2.2	2.1
427	20.5	9.2	5.2	32	0.06	2.6	0.7	1.9
188	15	20	5.8	43	0.13	4.8	1.7	1.6

۳-۵- تحليل رگرسيونی عيار طلا با رويکرد ترکیبی

رگرسیون تکنیکی است که سعی دارد روابط متغیرهای مختلف (در اینجا عناصر مختلف) را در قالب یک معادله ریاضی ساده نماید. در مطالعاتی که تعداد و نقش متغیرهای مستقل در تشکیل معادله رگرسیونی نامعلوم یا مبهم است، روش رگرسیون گام‌به‌گام می‌تواند مفید واقع شود. این روش به طور موثری مشخص می‌کند که چه ترکیبی از متغیرهای مستقل می‌تواند به بهترین وجه، متغیر وابسته را پیش‌بینی نماید [۲۸].

برای اجرای یک تحلیل رگرسیونی گام‌به‌گام سه الگوریتم پیشرو، پسرو و تلفیقی وجود دارند. در این پژوهش، الگوریتم تلفیقی به کار گرفته شده است چرا که ناتوانی الگوریتم‌های پیشرو و پسرو در اصلاح تصمیمات و حذف و اضافه کردن متغیرها را جبران می‌نماید. این روش برای هر متغیر، آماره F (آماره معنی‌داری الگوی رگرسیونی) را تشکیل می‌دهد. در این پژوهش مقدار بحرانی F برای ورود و خروج یک متغیر مستقل به معادله رگرسیونی برابر با مقدار ۰/۰۵ در نظر گرفته شده است. این معیار، کوچک در نظر گرفته شد تا حذف و اضافه متغیرها از مدل نهایی به طور نسبی آسان انجام شده و مدلی بهینه به دست آید.

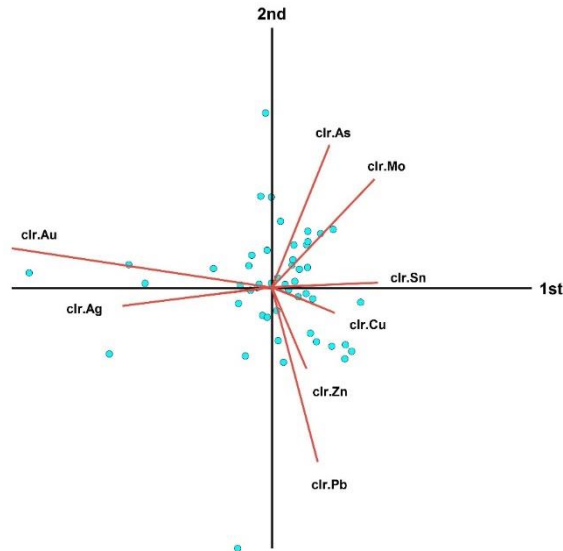
هدف از این پژوهش بررسی تاثیر ماهیت ترکیبی داده‌ها بر تحلیل رگرسیونی داده‌های لیتوژئوشیمیایی است. بدین منظور دو رویکرد متفاوت برای تحلیل رگرسیونی مقادیر آنومال اجرا و نتایج این دو رویکرد با یکدیگر مقایسه شدند. این دو رویکرد عبارتند از:

- رویکرد غیرترکیبی (سنتی) که در مطالعات ژئوشیمیایی متداول است و از اساس، مقادیر خام را تجزیه و تحلیل می‌نماید.

- رویکرد ترکیبی که ماهیت ترکیبی داده‌ها را در تجزیه و تحلیل تاثیر می‌دهد.

برای اجرای رویکرد ترکیبی نیاز است که داده‌ها باز شده و در محیطی که کوواریانس‌ها تحمیلی و دروغین نیستند بالانس‌ها تولید شوند. این کار با تعریف یک ماتریس باینری متوالی (Sequential Binary Partition (SBP)) میسر می‌شود. پیش از تعریف ماتریس SBP لازم است رفتار متغیرها در فضای باز نسبت به یکدیگر تحلیل شود. بدین منظور ابتدا متغیرها با تبدیل clr به فضای باز برده شدند و پس از محاسبه مولفه‌های اصلی (Principal Components)، رفتار متغیرها در نمودار دوتایی PC1 و PC2 تجزیه و تحلیل شد (شکل‌های ۵ و ۶).

بر اساس این دو نمودار، مشخص است که رفتار آماری طلا با سایر عناصر متفاوت است (شعاع $\text{clr}(\text{Au})$ بلندتر و در جهت مخالف سایر عناصر است). بدین ترتیب، در ماتریس SBP طلا به عنوان اولین جزء غیر هم‌رفتار از سایر عناصر جدا شده و نشانگر ۱- برای آن در نظر گرفته می‌شود. تکرار این روند برای دو جامعه آنومال منجر به تعریف دو ماتریس SBP به صورت رابطه (۱) می‌شود:



شکل ۶: نمودار دوتایی PC1-PC2 برای ۴۹ نمونه آنومال بر اساس حد جدایش $\bar{x} + 2s$ ، پس از باز کردن متغیرها با تبدیل clr. طلا با بلندترین شعاع، بیش از سایر عناصر در پوشش واریانس نقش دارد. پس از طلا، سرب و نقره نقش بارزی در پوشش واریانس دارند.

بر اساس ماتریس‌های SBP تعریف شده، بالانس‌های تبدیل ilr با رابطه (۲) محاسبه می‌شوند:

$$ilr_i = \sqrt{\frac{r_i + s_i}{r_i s_i}} \ln \left[\frac{g(z_-)}{g(z_+)} \right] \quad (2)$$

$i = 1, 2, \dots, D - 1,$

که در آن r_i تعداد مقادیر +۱ در ردیف i ام ماتریس SBP، s_i تعداد مقادیر -۱ در همان ردیف، $g(z_-)$ و $g(z_+)$ میانگین هندسی متغیرهایی هستند که در ماتریس SBP به ترتیب مقادیر +۱ و -۱ به آنها اختصاص داده شده است. بر اساس رابطه ۲ از ۸ متغیر عباری اولیه، ۷ بالانس تولید خواهد شد.

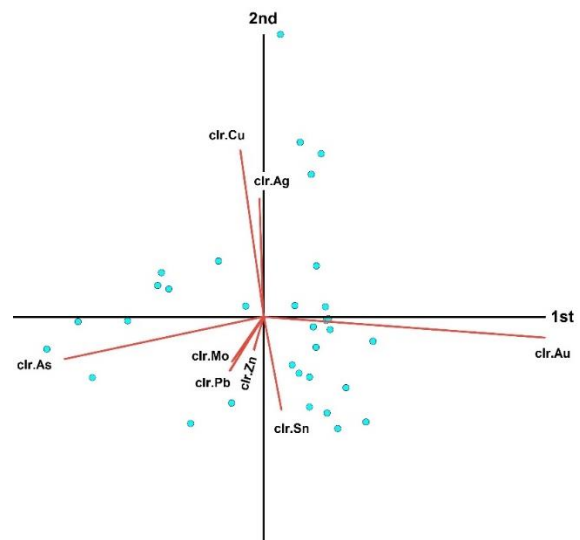
بر اساس ماتریس‌های SBP تعریف شده، طلا در بالانس اول از سایر متغیرها جدا شده است؛ لذا می‌توان با در نظر گرفتن بالانس اول به عنوان متغیر وابسته، آن را بر اساس سایر بالانس‌ها (که متغیر مستقل در نظر گرفته شده‌اند) با معادله رگرسیونی تخمین زد. این کار با روش رگرسیون گام به گام تلفیقی برای دو جامعه آنومال انجام شد و رابطه‌های رگرسیونی (۳) به دست آمدند.

$$y_1 = a + b.x_1 + c.x_2 + d.x_1^2 + e.x_2^2 + f.x_1.x_2 + g.x_1^3 + h.x_2^3 + i.x_1.x_2^2 + j.x_1^2.x_2, \quad (3)$$

که در آن y_1 مقادیر تخمینی ilr_1 ، x_1 معرف ilr_2 و x_2 معرف ilr_6 هستند. این معادله برای جامعه آنومال بر اساس

$$SBP_{(32 \text{ samples})} = \begin{bmatrix} Au & Cu & Pb & Zn & Ag & Sn & Mo & As \\ ilr_1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 \\ ilr_2 & 0 & +1 & +1 & +1 & +1 & +1 & -1 \\ ilr_3 & 0 & -1 & +1 & +1 & +1 & +1 & 0 \\ ilr_4 & 0 & 0 & +1 & +1 & -1 & +1 & 0 \\ ilr_5 & 0 & 0 & +1 & 0 & -1 & +1 & 0 \\ ilr_6 & 0 & 0 & -1 & +1 & 0 & 0 & +1 \\ ilr_7 & 0 & 0 & 0 & +1 & 0 & 0 & -1 \end{bmatrix},$$

$$SBP_{(49 \text{ samples})} = \begin{bmatrix} Au & Cu & Pb & Zn & Ag & Sn & Mo & As \\ ilr_1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 \\ ilr_2 & 0 & +1 & -1 & +1 & +1 & +1 & +1 \\ ilr_3 & 0 & +1 & 0 & +1 & -1 & +1 & +1 \\ ilr_4 & 0 & +1 & 0 & +1 & 0 & +1 & -1 \\ ilr_5 & 0 & +1 & 0 & +1 & 0 & +1 & -1 \\ ilr_6 & 0 & +1 & 0 & +1 & 0 & -1 & 0 \\ ilr_7 & 0 & +1 & 0 & -1 & 0 & 0 & 0 \end{bmatrix}. \quad (1)$$



شکل ۵: نمودار دوتایی PC1-PC2 برای ۳۲ نمونه آنومال بر اساس فواصل ماهالانوبیس. پس از باز کردن متغیرها با تبدیل clr. طلا به عنوان یک جزء منفرد با شعاع بلند و غیرهم‌جهت با سایر عناصر مشخص است. پس از طلا، نقش آرسنیک، مس و نقره در پوشش واریانس بارز است.

که ilr ها نشان دهنده بالانس‌ها، مقدار +۱ معرف رفتار همسان عناصر، مقدار -۱ معرف رفتار ناهمسان متغیرها و مقدار صفر معرف رفتار خنثی است.

معادله رگرسیونی تخمین طلا با رویکرد سنتی برای جامعه آنومال بر اساس حد جدایش $\bar{x} + 2s$ به صورت رابطه (۵) است:

$$y_3 = a + b.x_1 + c.x_2 + d.x_1^2 + e.x_2^2 + f.x_1.x_2 + g.x_1^3 + h.x_2^3 + i.x_1.x_2^2 + j.x_1^2.x_2, \quad (5)$$

که در آن y_3 مقادیر تخمینی طلا، x_1 معرف عیار عنصر روی و x_2 معرف عیار عنصر آرسنیک می‌باشند. معادله رگرسیونی جامعه آنومال با روش فواصل ماحالانوبیس به صورت رابطه (۶) است:

$$y_4 = a + \frac{b}{x_1} + \frac{c}{x_1^2} + \frac{d}{x_1^3} + \frac{e}{x_1^4} + \frac{f}{x_1^5} + g.\log(x_2) + h.\log(x_2^2) + i.\log(x_2^3) + j.\log(x_2^4) + k.\log(x_2^5), \quad (6)$$

که در آن y_4 مقادیر تخمینی طلا، x_1 معرف عیار عنصر آرسنیک و x_2 معرف عیار عنصر سرب هستند. مقادیر ثابت در معادله‌های بالا در جدول ۳ آورده شده‌اند. پارامترها و ضرایب معادلات رگرسیونی ۳ تا ۶ بر اساس بهینه‌سازی مقدار ضریب همبستگی داده‌های واقعی و تخمینی به دست آمده‌اند.

جدول ۳: ضرایب ثابت در معادلات رگرسیونی ۳ تا ۶

constant	y_1	y_2	y_3	y_4
a	-26.3804572042111	-413.516268381546	69.1071209199265	-4.45004142286879
b	-11.0105008090005	-1.14363700052018	-0.391019140844129	1.72075553668252
c	-35.2725266967386	1.61659954877865	-34.0724934411869	-8.37251088720984
d	3.56232148460984	0.380446485662727	-5.0510127893578E-04	14.8437174109386
e	-10.2504718263324	-0.273046264404076	5.07252481770976	-10.0551198777315
f	-0.861087218652294	-0.077280438755751	0.409848577792702	1.9762171119256
g	-0.178939972397824	6489.81791475879	-7.25505167784063E-06	7.5562633045676
h	-0.865644304794371	-38809.0833078877	-0.114697854822346	-4.90357466088094
i	0.324676648739061	111441.913836064	-5.63408434488149E-02	1.50988216804431
j	0.612007077151595	-153748.996047538	1.56881730314688E-03	-0.221215916927985
k	—	81790.5389693602	—	1.23729641405099E-02

ترکیبی در تخمین طلا به تفکیک روش‌های جدایش آنومالی مورد اعتبارسنجی و بحث قرار خواهند گرفت. به منظور اعتبارسنجی از معیارهای زیر استفاده شده است:

- نمودار پراکنش مقادیر واقعی و تخمینی
- نرخ طبقه‌بندی صحیح (Correct Classification Rate (CCR) که هرچه بزرگ‌تر باشد نشانه

حد جدایش $\bar{x} + 2s$ به دست آمده است. معادله رگرسیونی جامعه آنومال با روش فواصل ماحالانوبیس به صورت رابطه (۴) است:

$$y_2 = a + b.x_1 + c.x_1^2 + d.x_1^3 + e.x_1^4 + f.x_1^5 + \frac{g}{x_2} + \frac{h}{x_2^2} + \frac{i}{x_2^3} + \frac{j}{x_2^4} + \frac{k}{x_2^5}, \quad (4)$$

که در آن y_2 مقادیر تخمینی ilr_1 ، x_1 معرف ilr_2 و x_2 معرف ilr_3 هستند.

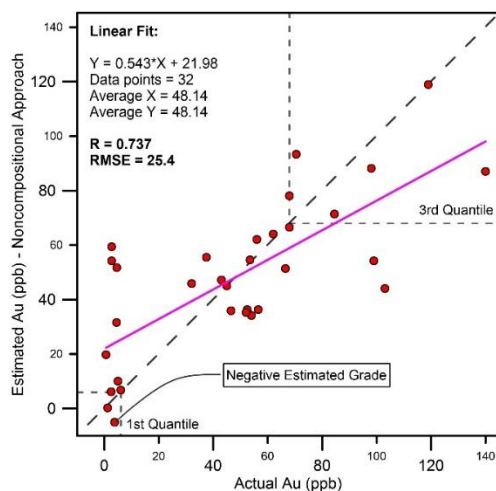
مشخص است که در تخمین بالانس اول در رابطه (۳)، دو بالانس ilr_2 و ilr_6 در رابطه (۴)، دو بالانس ilr_2 و ilr_3 دخیل بوده و سایر بالانس‌ها در معادلات رگرسیونی نهایی نقش کم اهمیت‌تری داشته و متعاقباً توسط الگوریتم تلفیقی از مدل رگرسیونی کنار گذاشته شده‌اند. پس از انجام تحلیل رگرسیونی در فضای باز، مقادیر تخمین زده شده برای بالانس اول به فضای بسته تبدیل معکوس شده و مقادیر طلای تخمینی بازیابی شدند. این کار امکان مقایسه مقادیر تخمینی و واقعی طلا را امکان‌پذیر می‌کند.

به منظور اجرای رویکرد سنتی، عیار طلا به عنوان متغیر وابسته در نظر گرفته شد. تحلیل رگرسیونی گام به گام تلفیقی نشان داد که عیار عناصر آرسنیک، سرب و روی در تخمین عیار طلا بیش‌ترین نقش را دارند.

۶- نتایج و بحث

به منظور مقایسه عملکرد رویکرد اول (سنتی) و دوم (ترکیبی)، تخمین‌های به دست آمده برای جامعه‌های آنومال تفکیک شدند. در ادامه، نتایج رویکرد سنتی و

است ضریب همبستگی مقادیر واقعی و تخمینی در رویکرد ترکیبی بزرگ‌تر از رویکرد سنتی است (۰/۸۳۶ در مقابل ۰/۷۳۷). به علاوه خطای جذر میانگین مربعات (Root Mean Squared Error (RMSE) برای رویکرد ترکیبی کمتر از رویکرد سنتی است (۲۰/۹ در مقابل ۲۵/۴). عطف به این دو معیار، رویکرد ترکیبی با دقت بیش‌تری طلا را تخمین زده است.

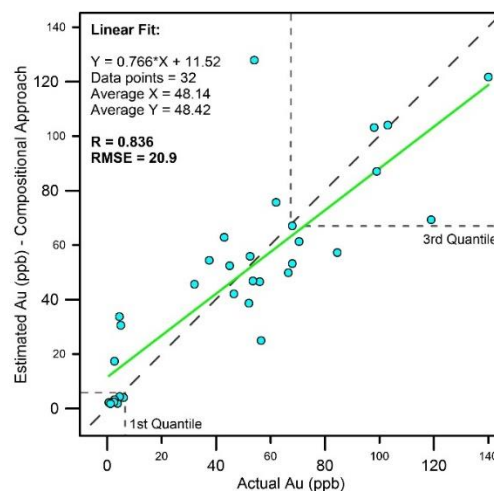


تخمین دقیق‌تر و در نتیجه، طبقه‌بندی عیاری بهتر است.

- مقایسه تابع توزیع مقادیر تخمینی با تابع توزیع مقادیر واقعی.

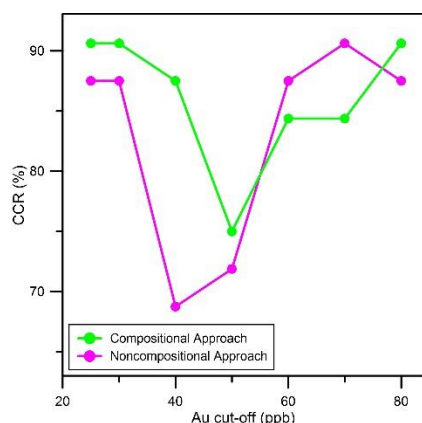
۶-۱- اعتبارسنجی تخمین طلا (جامعه آنومال بر اساس فواصل ماهالانویس)

نمودار پراکنش مقادیر واقعی و تخمینی به تفکیک دو رویکرد در شکل ۷ آورده شده است. همان طور که مشخص



شکل ۷: نمودار پراکنش مقادیر واقعی و تخمینی برای جامعه آنومال بر اساس فواصل ماهالانویس (راست: رویکرد ترکیبی؛ چپ: رویکرد سنتی)

ترکیبی طبقه‌بندی صحیح‌تری ارائه می‌نماید به این معنی که این رویکرد در تشخیص تعلق یک نمونه به کلاس عیاری تعریف شده، از رویکرد سنتی دقیق‌تر عمل می‌نماید (شکل ۸).



شکل ۸: مقایسه نرخ طبقه‌بندی صحیح دو رویکرد برای جامعه آنومال بر اساس فواصل ماهالانویس

این موضوع ریشه در بازتولید دقیق‌تر تابع توزیع مقادیر واقعی توسط رویکرد ترکیبی دارد (شکل‌های ۹ و ۱۰). در

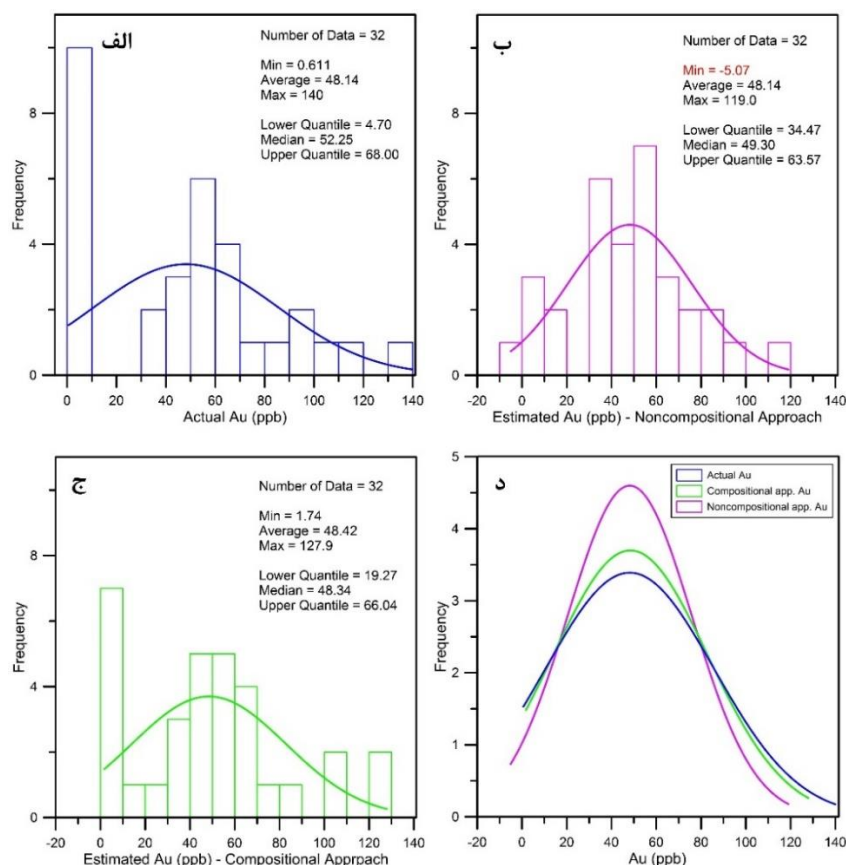
نکته قابل توجه، تخمین منفی عیار طلا توسط رویکرد سنتی در یک نقطه است. این تناقض آشکار نشان دهنده ضعف ذاتی روش‌های سنتی در تخمین داده‌های ترکیبی با سیستم بسته می‌باشد. در مقابل، رویکرد ترکیبی به گونه‌ای مقادیر طلا را بازتولید می‌کند که نتایج عیاری از تناقضات و مشکلات این‌چنینی باشند.

علاوه بر دقت و صحت عمومی تخمین‌ها، مطالعه عیار طلا در چارک‌های اول و سوم نشان می‌دهد که رویکرد ترکیبی در برآورد مقادیر این دو چارک از رویکرد سنتی دقیق‌تر عمل کرده است. در مقابل، رویکرد سنتی در بازتولید مقدار میانگین دقیق‌تر عمل کرده است (شکل ۷). از آنجایی که مقادیر قرار گرفته در چارک‌های اول و سوم نسبت به میانگین حیاتی‌تر هستند، تخمین دقیق آنها از اهمیت ویژه‌ای برخوردار است. عطف به این موضوع، مجدداً رویکرد ترکیبی تخمین‌های مطمئن‌تری ارائه نموده است.

مقایسه نرخ طبقه‌بندی صحیح (CCR) حاکی از آن است که در بیش‌تر عیار حدهای تعریف شده، رویکرد

مجموع نکات بالا حاکی از آن است که رویکرد ترکیبی در تخمین عیار طلا نسبت به رویکرد سنتی به میزان قابل ملاحظه‌ای دقیق‌تر عمل کرده و همچنین از تناقض‌هایی مانند تخمین عیارهای منفی دور بوده است. این موضوع، دلیلی بر برتری رویکرد ترکیبی و اهمیت روشی است که ماهیت ترکیبی داده‌های لیتوژئوشیمیایی را لحاظ می‌کند.

شکل‌های ۹ و ۱۰ هیستوگرام و نمودار احتمال جوامع حاصل از تخمین با جامعه مقادیر واقعی طلا مقایسه شده است. همان‌طور که مشخص است هیستوگرام رویکرد ترکیبی به مقادیر واقعی نزدیک‌تر است و این موضوع در خصوص نمودار احتمال (به خصوص در ابتدا و انتهای نمودار) نیز صدق می‌کند.

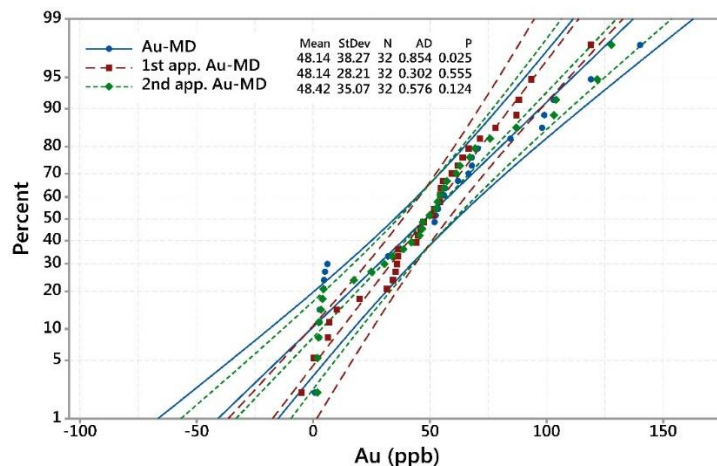


شکل ۹: مقایسه هیستوگرام مقادیر تخمینی با مقادیر واقعی طلا برای جامعه آنومال بر اساس فواصل ماهالانوبیس الف: مقادیر واقعی ب: مقادیر تخمینی با رویکرد غیر ترکیبی ج: مقادیر تخمینی با رویکرد ترکیبی د: هیستوگرام مقادیر تخمینی با رویکرد ترکیبی به هیستوگرام مقادیر واقعی طلا شبیه‌تر است

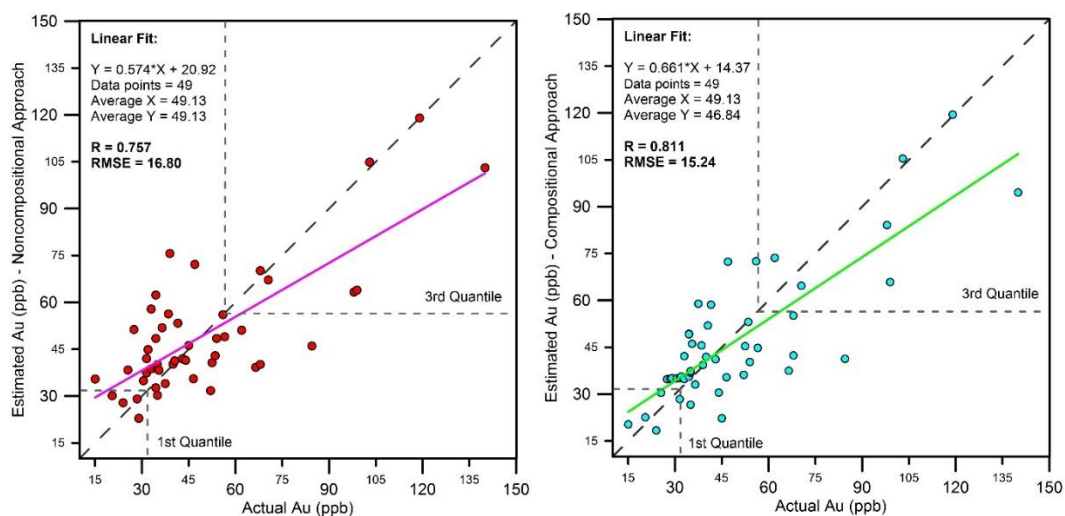
مقابل $(16/8)$. با توجه به این دو معیار، رویکرد ترکیبی با دقت بیشتری طلا را تخمین زده است. علاوه بر این، پراکندگی مقادیر تخمینی در چارک اول و سوم با رویکرد ترکیبی از رویکرد سنتی کم‌تر است و مقادیر تخمینی با رویکرد سنتی از خط $y = x$ فاصله بیشتری دارند. این موضوع بر دقت بیش‌تر تخمین با رویکرد ترکیبی در چارک‌های اول و سوم دلالت دارد. مقایسه نرخ طبقه‌بندی صحیح (CCR) نشان می‌دهد که تقریباً در تمامی عیار حدهای تعریف شده، رویکرد ترکیبی طبقه‌بندی صحیح‌تری ارائه می‌نماید (شکل ۱۲).

۲-۶- اعتبارسنجی تخمین طلا (جامعه آنومال بر اساس حد جدایش $(\bar{x} + 2s)$)

به مانند آنچه که در بخش ۱-۶ مطرح شد، نمودار پراکنش مقادیر واقعی و تخمینی با دو رویکرد سنتی و ترکیبی برای داده‌های آنومال بر اساس حد جدایش $\bar{x} + 2s$ مقایسه شدند. همان‌طور که در شکل ۱۱ مشخص است ضریب همبستگی مقادیر واقعی و تخمینی در رویکرد ترکیبی بزرگ‌تر از رویکرد سنتی است ($0/811$ در مقابل $0/757$). به علاوه خطای جذر میانگین مربعات برای رویکرد ترکیبی کم‌تر از رویکرد سنتی است ($15/2$ در



شکل ۱۰: مقایسه نمودار احتمال مقادیر تخمینی با مقادیر واقعی برای جامعه آنومال بر اساس فواصل ماحالانوبیس

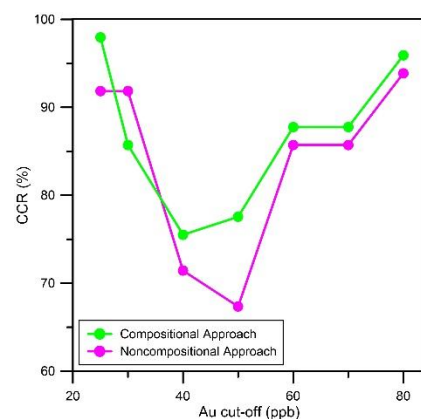


شکل ۱۱: نمودار پراکنش مقادیر واقعی و تخمینی برای جامعه آنومال بر اساس حد جدایش $\bar{x} + 2s$ (راست: رویکرد ترکیبی؛ چپ: رویکرد سنتی)

رویکرد ترکیبی در تخمین مقادیر بحرانی و بازتولید بهتر مقادیر واقعی طلا دلالت دارند.

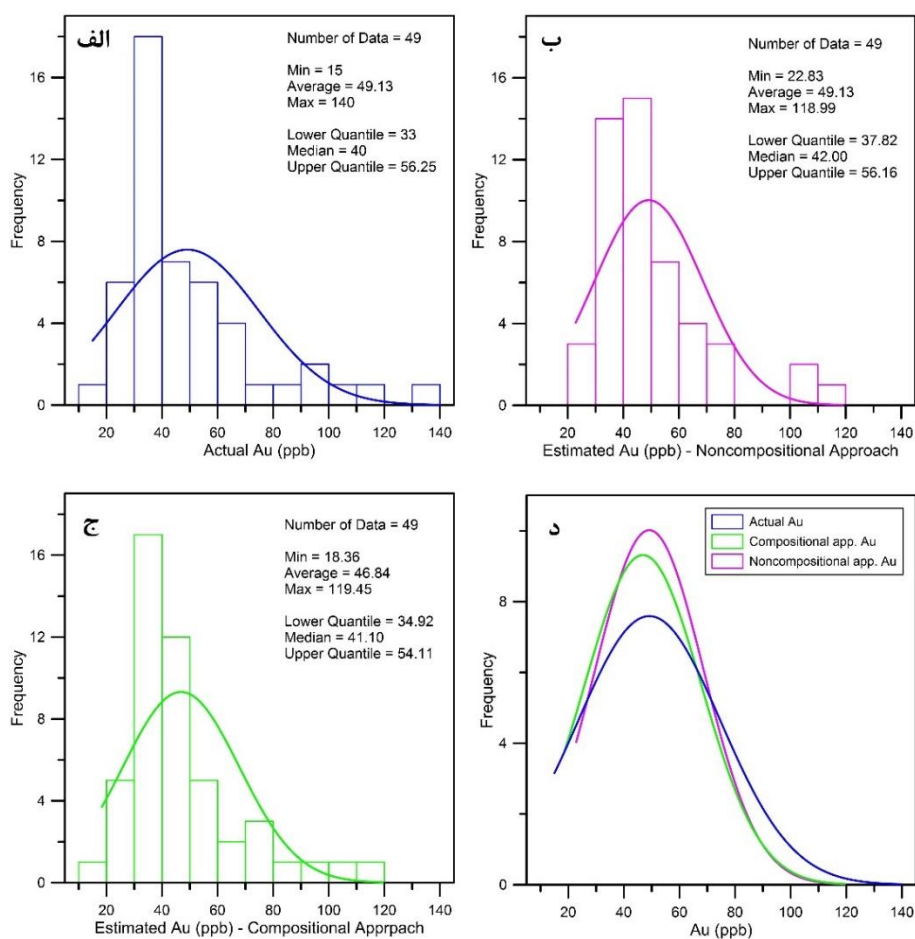
۷- نتیجه‌گیری

هدف از این پژوهش بررسی تاثیر ماهیت ترکیبی داده‌ها بر تحلیل رگرسیونی چندمتغیره و پاسخ به این سوال است که آیا در نظر گرفتن ماهیت ترکیبی داده‌ها در یک تحلیل رگرسیونی چندمتغیره لازم و مفید است یا خیر؟ بدین منظور، دو رویکرد سنتی و ترکیبی برای تحلیل رگرسیونی عیار طلا در یک رخداد معدنی طلا در شمال غرب ایران مورد بررسی قرار گرفتند و نتایج آن دو از نظر دقت و صحت مقایسه شد.

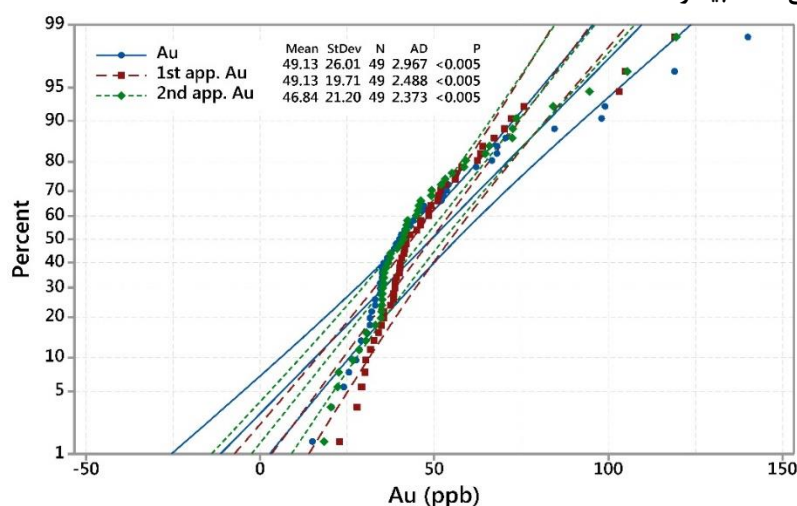


شکل ۱۲: مقایسه نرخ طبقه‌بندی صحیح دو رویکرد برای جامعه آنومال بر اساس حد جدایش $\bar{x} + 2s$

همچنین، این رویکرد، تابع توزیع مقادیر واقعی عیار طلا را بهتر از رویکرد سنتی بازتولید کرده است (شکل‌های ۱۳ و ۱۴). مجموع این نتایج بر دقت و صحت بیشتر



شکل ۱۳: مقایسه هیستوگرام مقادیر تخمینی با مقادیر واقعی طلا برای جامعه آنومال بر اساس حد جدایش $\bar{x} + 2s$ الف: مقادیر واقعی ب: مقادیر تخمینی با رویکرد غیر ترکیبی ج: مقادیر تخمینی با رویکرد ترکیبی د: هیستوگرام مقادیر تخمینی با رویکرد ترکیبی به هیستوگرام مقادیر واقعی طلا شبیه‌تر است



شکل ۱۴: مقایسه نمودار احتمال مقادیر تخمینی با مقادیر واقعی طلا برای جامعه آنومال بر اساس حد جدایش $\bar{x} + 2s$.

است. به صورت جزئی‌تر، می‌توان این گونه نتیجه‌گیری کرد که:

- اجرای رویکرد سنتی اگرچه سریع‌تر است و به دانش ریاضی کم‌تری نیاز دارد اما نتایج آن مستعد تناقض

با در نظر گرفتن معیارهای مختلف برای ارزیابی قابلیت اطمینان و دقت تخمین، مشخص شد که دقت و صحت رویکرد ترکیبی در تمامی شرایط از رویکرد سنتی بیشتر

Analysis in the Geosciences: from Theory to Practice. The Geological Society of London, 212 pp.

[2] Pawlowsky-Glahn, V., Egozcue, J. J., Lovell, D. (2015). Tools for compositional data with a total. *Stat. Modelling*, 15, 175–190.

[3] Pawlowsky-Glahn, V., Egozcue, J. J., Olea, R. A., Pardo-Igúzquiza, E. (2015). Cokriging of compositional balances including a dimension reduction and retrieval of original units. *Journal of South African Institution of Mining and Metallurgy*, 115.

[4] Aitchison J. (1984). The statistical analysis of geochemical compositions. *Mathematical Geology*, 16 (6), 531-564.

[5] Reimann, C., Filzmoser, P., Garrett, R., Dutter, G. (2008). *Statistical Data Analysis Explained: Applied Environmental Statistics with R*. 362 pp.

[6] Pawlowsky-Glahn, V., Egozcue, J. J. (2006). Compositional data and their analysis: an introduction. Geological Society, London, Special Publications, 264; 1-10.

[7] Chayes F. (1960). On correlation between variables of constant sum. *Journal of Geophysical Research*, 65 (12), 4185-4193.

[8] Chayes F. (1983). Detecting nonrandom associations between proportions by tests of remaining space variables. *Mathematical Geology*, 15 (1), 197- 206.

[9] Chayes F. (1971). *Ratio Correlation*. University of Chicago Press, Chicago, IL, 99 pp.

[10] Butler, J. C. (1976). Principal components analysis using the hypothetical closed array. *Mathematical Geology*, 8, 25–36.

[11] Darroch, J.N. (1969). Null correlations for proportions. *Mathematical Geology*, 1, 221–227.

[12] Pawlowsky-Glahn, V. (1984). On spurious spatial covariance between variables of constant sum. *Sci. Terre Sér. Inform*, 21, 107–113.

[13] Carranza J. (2006). Analysis and mapping of stream sediment geochemical anomalies: should we logratio transform the data?, University of Twente.

[14] Pawlowsky-Glahn, V., Olea, R. A., Davis, J.C. (1995). Estimation of regionalized compositions: a comparison of three methods. *Math. Geol*, 27 (1).

[15] Pawlowsky-Glahn, V., Olea, R. A. (2004). Geostatistical analysis of compositional data. In: Number 7 in *Studies in Mathematical Geology*. University Press, Oxford.

هستند. به عنوان مثال، در این پژوهش رویکرد سنتی یک مقدار منفی برای عیار طلا برآورد نمود که به وضوح، یک تناقض ریاضی آشکار است.

- رویکرد ترکیبی ارائه شده در این پژوهش ماهیت ترکیبی داده‌ها را در نظر می‌گیرد و با حذف کوواریانس دروغین و همبستگی تحمیلی میان متغیرها نتایجی عاری از تناقض‌هایی مانند تخمین عیارهای منفی ارائه می‌کند.

- تابع توزیع مقادیر تخمینی با رویکرد ترکیبی، بیش از رویکرد سنتی به مقادیر واقعی طلا شبیه است که منجر به طبقه‌بندی واقعی‌تر نمونه‌ها می‌شود. نتیجتاً، نرخ طبقه‌بندی صحیح (CCR) رویکرد ترکیبی بزرگ‌تر از رویکرد سنتی است.

- اگرچه رویکرد سنتی در برآورد میانگین جامعه دقیق‌تر از رویکرد ترکیبی عمل می‌کند اما دقت رویکرد ترکیبی در تخمین مقادیر طلا در چارک‌های اول و سوم به صورت معناداری از رویکرد سنتی بیشتر است. با توجه به اینکه مقادیر قرار گرفته در این دو چارک برای ارزیابی‌های اکتشافی حیاتی‌تر هستند لذا رویکرد ترکیبی گزینه مطمئن‌تری برای اکتشاف ذخایر مشابه به حساب می‌آید.

با اینکه در اغلب مطالعات اکتشافی کشور، تحلیل داده‌های ترکیبی نادیده گرفته می‌شود، مجموع نکات بالا نشان‌دهنده ضرورت در نظر گرفتن ماهیت ترکیبی داده‌ها در مطالعات آماری داده‌های ژئوشیمیایی هستند. با توجه به قابلیت اطمینان و دقت بیشتر رویکرد ترکیبی ارائه شده در این پژوهش، استفاده از رویکردی مشابه در سایر مطالعات و پروژه‌های اکتشافی پیشنهاد می‌شود.

۸- تقدیر و تشکر

بدین وسیله از پروفیسور ورا پاولوفسکی-گلن و پروفیسور خوان خوزه اگوزکیو به خاطر راهنمایی‌های ارزشمند و کمک علمی در روند پیشبرد این پژوهش قدردانی و تشکر به عمل می‌آید.

مراجع

[1] Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (2006). *Compositional Data*

- [27] Moeini, H., Torab, F. M. (2017). Comparing compositional multivariate outliers with autoencoder networks in anomaly detection at Hamich exploration area, east of Iran. *Journal of Geochemical Explorations*, 180, 15–23.
- [28] Takezawa, K. (2006). Introduction to nonparametric regression. John Wiley and sons, Inc., publications, 546 pages.
- [16] Aitchison J. (1986). The statistical analysis of compositional data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd, London. Reprinted, with additional material by The Blackburn Press, Caldwell, NJ.
- [17] Egozcue J. J., Pawlowsky-Glahn, V., Ateufigueras, G., Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35 (3), 279-300.
- [18] Molayemat, H., Torab, F. M., Pawlowsky-Glahn, V., Hossein Morshedy, A., Egozcue, J. J. (2018). The impact of the compositional nature of data on coal reserve evaluation, a case study in Parvadeh IV coal deposit, Central Iran. *International Journal of Coal Geology*, 188, 94-111.
- [19] Govvet, G. J. S. (1983). *Rock Geochemistry in Exploration*. Elsevier Scientific Publications Co. 461pp.
- [20] Kouhestani, H., Ghaderi, M., Chang, Z., Zaw, K. (2015). Constraint on the ore fluids in the Chah Zard breccia-hosted epithermal Au-Ag deposit, Iran: fluid inclusion and stable isotope studies. *Ore Geology Reviews*, 65, 512-521.
- [21] Kouhestani, H., Rastad, E., Rashidnezhad Omran, N., Mohjel, M. (2013). Controlling settings, genesis and forming model of Chah-bagh gold deposit, Muteh mine ditrict. *Geoscience Journal of Geology Survey of Iran*, Vol. 88.3. (in Persian).
- [22] Niroumand, S., Moor, F., Mohjel M., Emami, M. H., Mehrparto, M., Borna, B. (2013). Kharapeh gold deposit: an example of epi-zonal gold occurrence, NE of Piranshahr, Western Azarbayjan province, NW Iran. *Geoscience Journal of Geology Survey of Iran*, Vol. 88.3. (in Persian).
- [23] Ghavami Riabi, R., Shahsavani, D. (2012). Application of the MARS modeling to estimate the gold values based on the paragenetic elements for an exploration data set. *Iranian Journal of Mining Engineering*, 7(15), 15-23. (in Persian).
- [24] Hariri, A. (2003). Geological report of 1:100,000 geological map of Saqez. *Geological Survey of Iran*. (in Persian).
- [25] Novin Ekteshaf Tose'e consulting engineers, (2008). Exploration and prospecting of gold in the Qareh Char area, 62 pages.
- [26] Hosseini, S. A., Eftekhari Mahabadi, S., Asghari, O. (2015). The Comparison of Appropriate Methods in imputation of the censored values in the geochemical datasets. *Journal of Analytical and Numerical Methods in Mining Engineering* 5(9), 63-72 (in Persian).

¹ isometric logratio-Expectation Maximization (ilr-EM)