IJAE
Iranian Journal of Applied Ecology
بوم‌شناسی کاربردی/ سال چهاردهم / شماره چهارم / ۱٤۰٤/ صفحات ۷۱ تا ۸۹

ارزیابی توزیع آماری و روش‌های تبدیل داده درصد پوشش گیاهی در مراتع منطقه حفاظت‌شده درمیان–سربیشه

مسلم رستم پور*

(تاریخ دریافت: ۱٤۰٤/۱۱/۱۵ ؛ تاریخ پذیرش: ۱٤۰۵/۰٤/۱۷)

چکیده

در این پژوهش، توزیع آماری درصد پوشش گیاهی و روش‌های تبدیل داده در ۱۰ تیپ مرتعی منطقه حفاظت‌شده درمیان-سربیشه بررسی شد. پوشش گیاهی در ٦۰۰ پلات ٤ مترمربعی اندازه‌گیری گردید. نرمال بودن داده‌ها با ضریب چولگی، نمودارهای گرافیکی و آزمون‌های شاپیرو-ویلک، آندرسون-دارلینگ و لیلی‌فرز ارزیابی شد. نتایج نشان داد که به‌جز یک تیپ، داده‌های ۹ تیپ دیگر نرمال نبوده و دارای چولگی راست متوسط تا زیاد بودند. برازش چهار توزیع نرمال، لوگ‌نرمال، گاما و ویبول با معیار اطلاعات آکائیک (AIC) نشان داد که توزیع نرمال برای ۹۰ درصد تیپ‌ها نامناسب است؛ به‌طوری‌که لوگ‌نرمال در ۸۰ درصد و ویبول در ۲۰ درصد تیپ‌ها بهترین برازش را داشتند. از شش روش تبدیل داده (ریشه دوم، ریشه سوم، لگاریتم طبیعی و پایه ۱۰، معکوس و باکس-کاکس)، تنها تبدیل‌های لگاریتمی و باکس-کاکس در ۵۰ درصد تیپ‌ها موفق به نرمال‌سازی شدند و تبدیل معکوس کاملاً ناموفق بود. با توجه به نرمال نبودن ذاتی بیشتر داده‌های پوشش گیاهی در مراتع نیمه‌خشک و کارایی محدود تبدیل‌ها، استفاده از روش‌های مقاوم مانند تحلیل مؤلفه‌های اصلی مقاوم (Robust PCA) که وابستگی کمتری به پیش‌فرض نرمال بودن دارند، برای تحلیل‌های جامعه‌شناسی گیاهی پیشنهاد می‌شود. یافته‌های این پژوهش می‌تواند در انتخاب روش آماری مناسب برای مطالعات بوم‌شناختی و مدیریت مراتع راهگشا باشد.

واژه‌های کلیدی: آمار چندمتغیره، پوشش گیاهی، توزیع لوگ نرمال، نکویی برازش.

دانشیار گروه مرتع و آبخیزداری و عضوگروه پژوهشی آینده پژوهی در علوم محیطی، دانشکده منابع طبیعی و محیط زیست، دانشگاه بیرجند، بیرجند، ایران.
* مسئول مکاتبات، پست الکترونیکی: rostampour@birjand.ac.ir

مقدمه

درصد پوشش گیاهی، یکی از مهمترین خصوصیات اکوسیستم‌های مرتعی است که در طرح‌های پایش و ارزیابی مراتع اندازه‌گیری می‌شود. از داده‌های درصد پوشش گیاهی در طبقه‌بندی و رسته‌بندی جوامع گیاهی (۲۲، ۲۹، ٤۷ و ۵۱)، مطالعات روابط خاک و گیاه (۲۰)، برآورد تولید علوفه (۲ و ۱۰)، اندازه‌گیری ترکیب گیاهی (۳۱)، تنوع گونه‌ای (۲٤)، ارزش اهمیت نسبی (٤۰)، تعیین وضعیت مرتع (۳۲)، تحلیل عملکرد چشم‌انداز و سلامت مرتع (۳۳)، برآوردهای فرسایش و رسوب و شایستگی مرتع (۱ و ۳٤) استفاده می‌شود. در تحلیل‌های آماری مطالعات جامعه‌شناسی گیاهی، روابط خاک و گیاه و تحلیل‌های آماری در مطالعات ارزیابی مراتع مثل برآورد تولید علوفه از طریق رگرسیون، آگاهی از توزیع آماری داده‌های پوشش گیاهی ضرورت دارد. اصولاً در پژوهش‌های علوم مرتع که با آمار استنباطی سروکار داریم و قرار است آماره‌های نمونه به پارامترهای جامعه تعمیم داده شود، در این موارد فرض می‌شود داده‌ها، معرف جامعه هستند، از این‌رو آگاهی از توزیع آماری داده‌ها و بویژه فرض نرمال بودن داده‌ها ضروری است.

استفاده از آزمون‌های پارامتریک تک متغیره و دو متغیره مثل تحلیل واریانس (ANOVA) برای مقایسه پوشش گیاهی در تیپ‌های گیاهی مختلف و تحلیل رگرسیون برای برآورد تولید علوفه نیازمند داده‌هایی با توزیع نرمال است. آزمون تحلیل واریانس فرض می‌کند که متغیر وابسته بایستی نرمال بوده و نمونه‌ها باید از جمعیتی با واریانس همگن تهیه شوند (۵ و ۵۲). با این حال، شواهد تجربی مربوط به داده‌های واقعی استخراج شده از بررسی چندین مجله علمی نشان می‌دهد که این فرضیات همیشه برآورده نمی‌شوند (۸). نرمال بودن و فرضیات دیگر (همگنی واریانس‌ها و استقلال خطاها) باید جدی گرفته شود، زیرا وقتی این فرضیات تایید نشود، نتیجه‌گیری، دقیق و قابل اعتماد نخواهد بود (۱۵). بوم‌شناسان گیاهی که قصد دارند از آزمون‌های پارامتریک براساس توزیع نرمال استفاده کنند، نیاز به بررسی نرمال بودن داده‌ها دارند (۲٦).

روش‌های مختلفی برای بررسی نرمال بودن داده‌ها وجود دارد، اما در عموم پژوهش‌های علوم مرتع از روش‌های آماری کولموگروف-اسمیرنوف یا شاپیرو-ویلک استفاده می‌شود، اما بایستی اشاره کرد که این آزمون‌ها وابسته به اندازه نمونه (تعداد پلات) هستند، اگر اندازه نمونه کوچک باشد، این آزمون‌ها دارای توان آماری پایینی هستند، از طرفی ممکن است انحراف از نرمال بودن را حتی اگر بسیار کوچک باشد را معنی‌دار تشخیص دهد (٤۸). با این حال، استفاده از روش‌هایی مثل ترسیم نمودارهای هیستوگرام، چندک-چندک، جعبه‌ای و آماره‌های توصیفی مثل چولگی و کشیدگی از نظرها اغفال مانده است.

در منابع پایه بوم‌شناسی کمی و جامعه شناسی گیاهی اشاره شده است که برخی از داده‌های پوشش گیاهی با توزیع گسسته مثل تراکم گیاهی از توزیع دوجمله‌ای منفی یا پواسون برخوردار هستند (۲۲، ۲٤ و ۲٦)، یا Magurran و McGill (۲۰۱۱) اذعان می‌کند که داده‌های وفور گونه‌ای از توزیع لوگ نرمال تبعیت می‌کنند (۲۷). با مروری بر پژوهش‌های بوم شناسی گیاهی می‌توان دریافت که داده‌های مختلف پوشش گیاهی از توزیع‌های غیرنرمال تبعیت می‌کنند، به عنوان نمونه، Ulrich و همکاران (۲۰۲۲) در فراتحلیلی، داده‌های درصد پوشش گیاهی ۱۷۲۵ پلات ۱۰ متر مربعی از پایگاه داده GrassPlot شامل ٤۷ مجموعه داده منطقه‌ای از ۱۹ علفزار قلمرو جغرافیایی پالئارکتیک را بررسی کردند. نتایج آن‌ها نشان داد که داده‌های پوشش گیاهی از توزیع ویبول که مشابه توزیع‌های سری لگاریتمی و لوگ نرمال است تبعیت می‌کنند (٤۵). Callaghan و همکاران (۲۰۲۳) در فراتحلیلی، گزارش می‌کنند که شواهد قوی درسترس است که پوشش گیاهی از توزیع لوگ نرمال پواسون و دوجمله‌ای منفی تبعیت می‌کنند (۱٤). اکثر این‌گونه پژوهش‌ها بر روی داده‌های تراکم؛ غنا و وفور گونه‌ای است، با این حال، Kattge و همکاران (۲۰۱۱) در فراتحلیلی، با مطالعه ۵۲ گروه از صفات مورفولوژیکی، آناتومیکی، فیزیولوژیکی، بیوشیمیایی و فنولوژیکی گیاهان و اندام‌های آنها در ۹۳ پایگاه داده مربوط به ٦۹۰۰۰ گونه گیاهی در سرتاسر جهان اذعان کردند که اکثر این

خصوصیات تقریباً از توزیع لوگ نرمال تبعیت می‌کنند (۲۱). از این‌رو نرمال دانستن داده‌های پوشش گیاهی بدون بررسی توزیع آنها جای اشکال دارد. از نظر آماری، داده‌های بیولوژیک اغلب از توزیع نرمال برخوردار نیستند و تبدیل‌های متعددی برای بهبود نرمالیتی داده‌ها پیشنهاد می‌شود (۵۰). انتخاب روش تبدیل داده‌ها نیز اصول و معیاری دارد اما متاسفانه در پژوهش‌های علوم مرتع، بدون در نظر گرفتن آنها از تبدیل رایج لگاریتمی استفاده می‌شود.

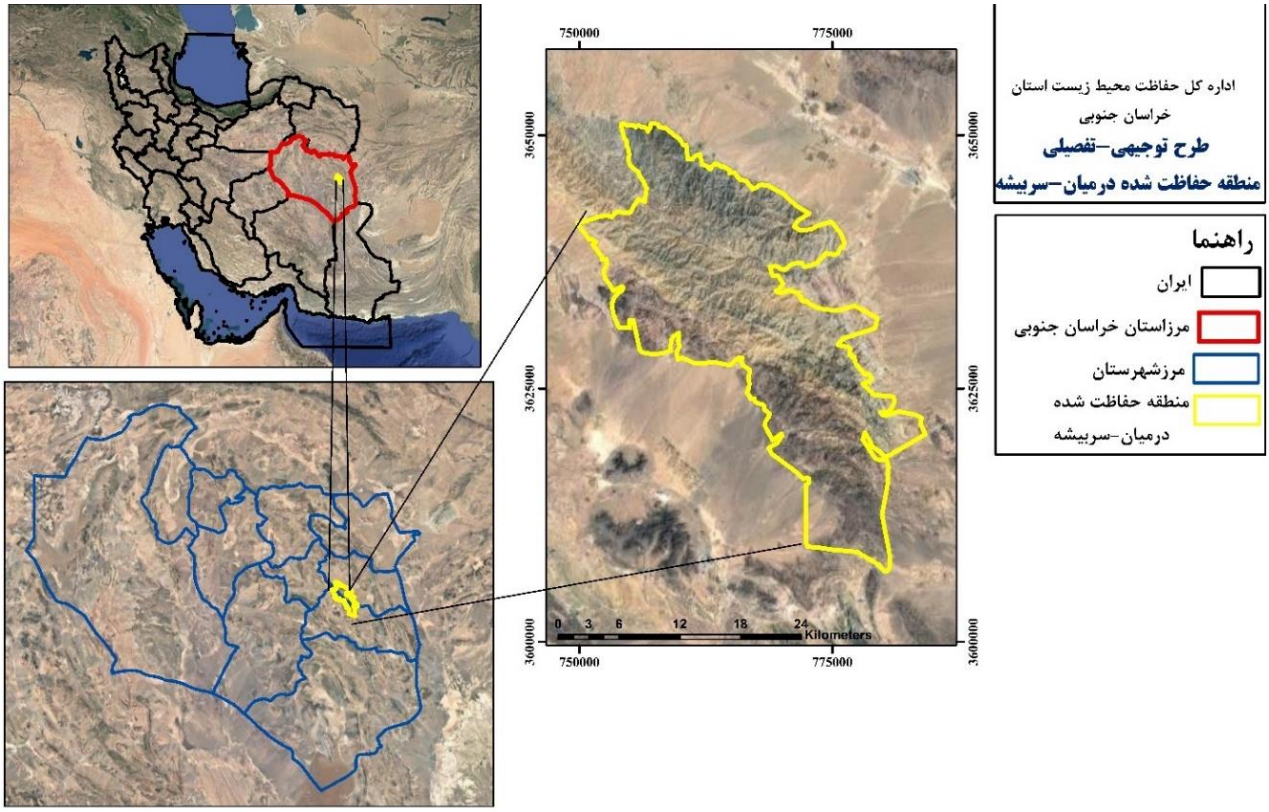
بررسی پیشینه پژوهش نشان می‌دهد که اگرچه توزیع آماری داده‌های تراکم، وفور، غنا و تنوع گونه‌ای در مطالعات متعدد بوم‌شناسی گیاهی مورد تحلیل قرار گرفته است، اما اطلاعات موجود در خصوص برازش داده‌های درصد پوشش گیاهی با توزیع‌های آماری پیوسته به ویژه در مراتع نیمه‌خشک ایران بسیار محدود است. همچنین، مطالعات جامعی که به مقایسه کارایی روش‌های مختلف تبدیل داده در نرمال‌سازی درصد پوشش گیاهی پرداخته باشد، به ندرت یافت می‌شود.

پژوهش حاضر با به‌کارگیری یک رویکرد تلفیقی و چندمرحله‌ای، برای نخستین بار در منطقه حفاظت‌شده درمیان–سربیشه به ارزیابی سیستماتیک توزیع آماری درصد پوشش گیاهی می‌پردازد. تمایز اساسی این مطالعه با پژوهش‌های پیشین در سه جنبه است: اول، به جای اکتفا به یک آزمون نرمالیتی (که معمولاً در مطالعات مرتع رایج است)، از ترکیب چهار گروه روش شامل آزمون‌های آماری (شاپیرو–ویلک، آندرسون–دارلینگ و لیلی‌فرز)، شاخص‌های توصیفی (چولگی و کشیدگی)، نمودارهای گرافیکی (هیستوگرام، جعبه‌ای و Q-Q plot) و برازش چهار توزیع پیوسته با معیار AIC استفاده می‌کند که خطای تشخیص الگوی توزیع را به حداقل می‌رساند. دوم، به جای بررسی صرفاً وجود یا عدم نرمال بودن، به شناسایی دقیق نوع توزیع (نرمال، لوگ‌نرمال، گاما یا ویبول) در هر تیپ گیاهی می‌پردازد و سوم، کارایی شش روش تبدیل داده را به صورت کمّی می‌سنجد و بر اساس شواهد تجربی، کاربرد روش‌های آماری مقاوم را جایگزین رویکردهای کلاسیک وابسته به نرمال بودن پیشنهاد می‌کند. این رویکرد ترکیبی می‌تواند به عنوان الگویی برای پژوهش‌های آینده در جامعه‌شناسی گیاهی مناطق نیمه‌خشک مورد استفاده قرار گیرد.

مواد و روش‌ها

مشخصات منطقه مورد مطالعه

منطقه حفاظت‌شده درمیان–سربیشه در بین دو شهرستان درمیان و سربیشه واقع شده است. شهرستان درمیان از شمال با شهرستان قائنات، از شرق با کشور افغانستان، از جنوب با شهرستان سربیشه و از غرب با شهرستان بیرجند هم‌مرز است. مرکز این شهرستان، شهر اسدیه، در موقعیت جغرافیایی حدود ٦٠ درجه و ۱ دقیقه طول شرقی و ۳۲ درجه و ۵۵ دقیقه عرض شمالی قرار دارد. بر اساس سرشماری سال ۱۳۹۰، شهر اسدیه نزدیک به شش هزار نفر جمعیت داشته است. اسدیه در فاصله‌ی حدود ۹۵ کیلومتری شرق بیرجند (مرکز استان خراسان جنوبی) واقع شده و ارتفاع آن از سطح دریا تقریباً ۱۵۰۰ متر است.شهرستان سربیشه نیز از شمال با شهرستان درمیان، از شرق با افغانستان، از جنوب با شهرستان نهبندان و از غرب با شهرستان بیرجند هم‌مرز است و مرکز آن شهر سربیشه است. به سبب قرار گرفتن در ناحیه مرزی، این دو شهرستان از اهمیت جغرافیایی و استراتژیک بالایی برخوردارند. طبیعت منطقه در بخش‌های کوهستانی شمالی و مرکزی، زیستگاه گونه‌های جانوری و گیاهی ارزشمند است، درحالی‌که در دشت‌های جنوبی و شرقی، شرایط نیمه‌خشک و بیابانی غالب است. ترکیب این شرایط طبیعی، همراه با حضور جمعیت‌های روستایی و عشایری، باعث شده است که منطقه حفاظت‌شده درمیان–سربیشه ازنظر تنوع زیستی و انسانی اهمیت ویژه‌ای داشته باشد. پوشش گیاهی منطقه عمدتاً از تیپ‌های بوته‌زار و مرتعی سازگار با اقلیم خشک تشکیل شده است و در برخی دره‌ها و کوهپایه‌های مرطوب‌تر، لکه‌هایی از درختچه‌های بومی مانند بنه و بادامک مشاهده می‌شود. این ترکیب ناهمواری، اقلیم خشک و پوشش گیاهی نیمه‌بیابانی، منظر طبیعی منطقه را به سیمایی زنده و چشم‌نواز تبدیل کرده است که



شکل ۱. موقعیت جغرافیایی منطقه حفاظت شده درمیان- سربیشه

پیوستگی بصری میان کوهستان، دامنه و دشت را حفظ می‌کند.منطقه حفاظت‌شده درمیان-سربیشه بر اساس مصوبه شماره ۷٤٤ شورای عالی محیط‌زیست مورخ ۱۳۹۷/۵/۱ به عنوان منطقه حفاظت‌شده ثبت و ابلاغ شده است (۳۹).

نمونه‌برداری از پوشش گیاهی

پژوهش حاضر در ۱۰ تیپ گیاهی منطقه حفاظت شده درمیان-سربیشه، استان خراسان جنوبی انجام شده است. نمونه-برداری در خرداد سال ۱٤۰۲ در منطقه معرف تیپ‌های گیاهی انجام شده است. تیپ‌های گیاهی قبلاً به روش فیزیونومیک-فلورستیک شناسایی و نامگذاری شده بود (۳۹). نمونه‌برداری از پوشش گیاهی به‌روش تصادفی-سیستماتیک، با استفاده از ترانسکت‌های ۲۰۰ متری و کوادرات در هر منطقه صورت گرفت. مؤثرترین شکل و اندازه کوادرات با استفاده از الگوی پراکنش و اندازه گیاهان منطقه، سطح پلات ۲×۲متری در نظر گرفته شد. در هر تیپ، سه ترانسکت در جهت عمود بر شیب و به‌صورت موازی مستقر شد. در هر ترانسکت با فاصله ۱۰ متر، یک کوادرات قرار داده شد (جمعاً تعداد ۲۰ کوادرات در هر ترانسکت و ٦۰ کوادرات در کل تیپ گیاهی و ٦۰۰ کوادرات در کل منطقه) و درصد پوشش گیاهی به تفکیک گونه‌های گیاهی اندازه‌گیری شد. ماتریس داده‌های گیاهی، دقیقاً همان ماتریسی بود که در تحلیل‌های چندمتغیره جامعه‌شناسی گیاهی و ارزیابی تنوع زیستی استفاده می‌شود؛ بدین صورت که تیپ‌های گیاهی در ستون‌ها، گونه‌های گیاهی در ردیف‌ها و در هر سلول نیز میانگین درصد پوشش گیاهی هر گونه وارد شد. از آنجایی که محاسبات و ترسیم نمودارها برای هر تیپ به صورت جداگانه انجام شد، داده‌ها به صورت وکتوری به محیط نرم افزاری R نسخه ٤.٦.۰ (۳۸) وارد شدند. بنابراین لازم به تأکید است که این شیوه ورود داده، برای تحلیل‌های آماری تک‌متغیره و دومتغیره در طرح‌های آزمایشی منابع طبیعی نظیر طرح‌های پایه (CRD،

جدول ۱. روش های عددی و تصویری سنجش نرمال بودن داده‌ها

ماهیت روش‌ها	انواع روش ها	
	روش‌های عددی	روش‌های گرافیکی
توصیفی	چولگی	نمودار هیستوگرام نمودار جعبه‌ای نمودار ساقه و برگ
مبتنی بر نظریه	آزمون شاپیرو- ویلک آزمون کلموگروف- اسمیرنوف آزمون آندرسون- دارلینگ آزمون رایان- جوینر آزمون جارک- برا	نمودار احتمال –احتمال (P-P) نمودار چارک –چارک (Q-Q)

RCBD و LS) و فاکتوریل مناسب نمی‌باشد، بلکه صرفاً متناسب با الگوی داده‌های مطالعات جامعه‌شناسی گیاهی در نرم‌افزارهایی همچون PC-ORD، CANOCO، PAST و R است.

تجزیه و تحلیل آماری

روش‌های سنجش نرمال بودن داده‌ها

روش‌های بررسی نرمالیتی به دو گروه اصلی تقسیم می‌شوند که عبارتند از روش‌های توصیفی و روش‌های استنباطی (آزمون-های آماری) (جدول ۱).

روش‌های توصیفی:

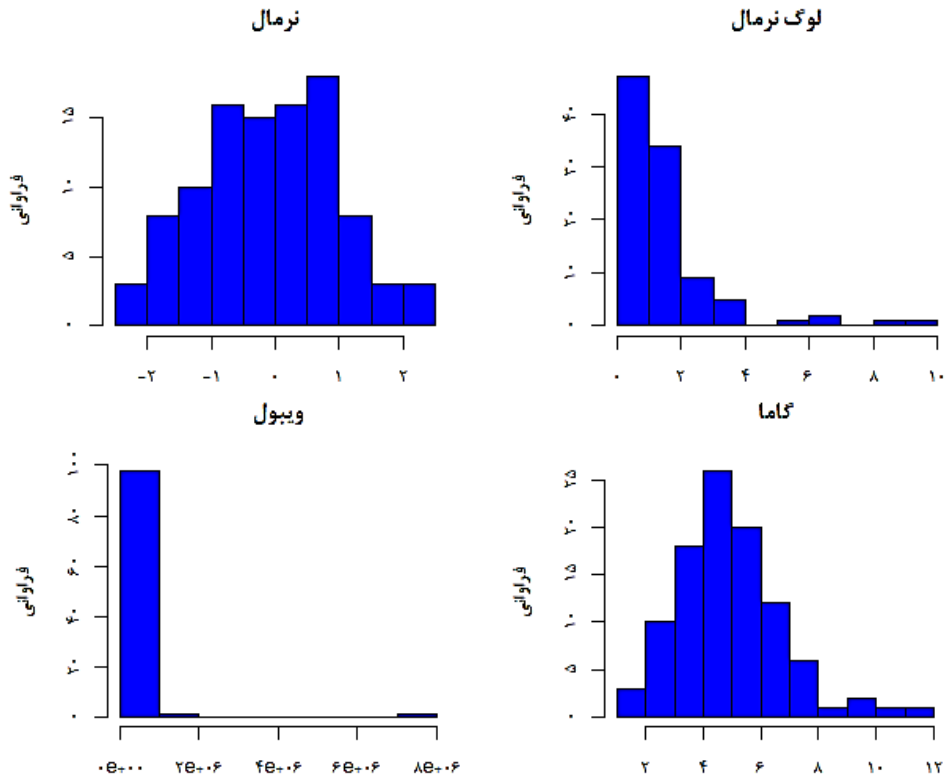
۱. ضریب چولگی (گشتاور نوع اول) به همراه خطای استاندارد آن برای داده‌های درصد پوشش گیاهی تیپ‌های گیاهی مورد مطالعه محاسبه شد. پس از تقسیم مقادیر چولگی بر خطای استلندارد آن،اگر مقدار Z در بازه ۲ و ۲- باشد داده‌ها نرمال است.

۲. نمودار هیستوگرام: نمودار توزیع فراوانی داده‌های (N= ٦٠) چهار توزیع نرمال، لوگ نرمال، گاما و ویبول ترسیم شد (شکل ۲). مطابقت داده‌های پوشش گیاهی منطقه مورد مطالعه با الگوی هر کدام از توزیع‌های مربوطه، حاکی از نرمال بودن/نبودن داده است.

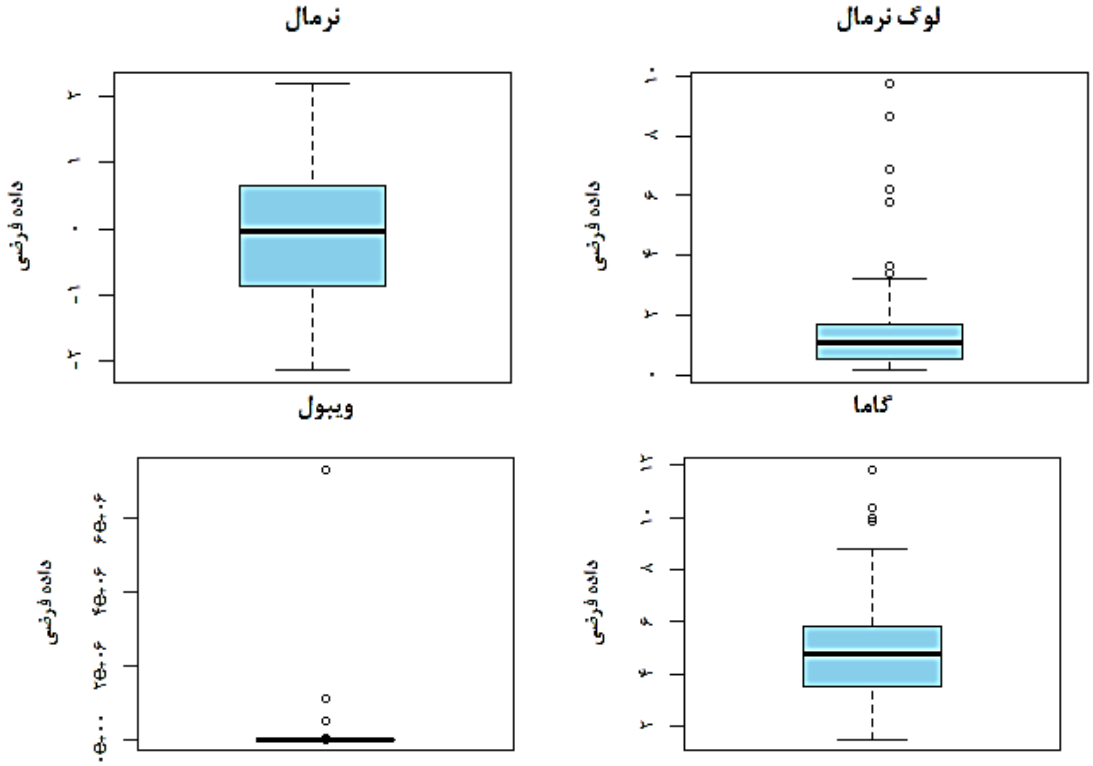
۳. نمودار جعبه‌ای یک روش استاندارد برای نمایش توزیع داده‌ها است که براساس شاخص‌های آماری حداقل ، چارک اول، میانه، چارک سوم و حداکثر ترسیم می‌شود. از این نمودار می‌توان داده(های) پرت را تشخیص داد. مشابه شکل ۳، در صورت عدم وجود داده پرت، نرمال بودن داده‌ها تایید می‌شود.

روش‌های مبتنی بر نظریه

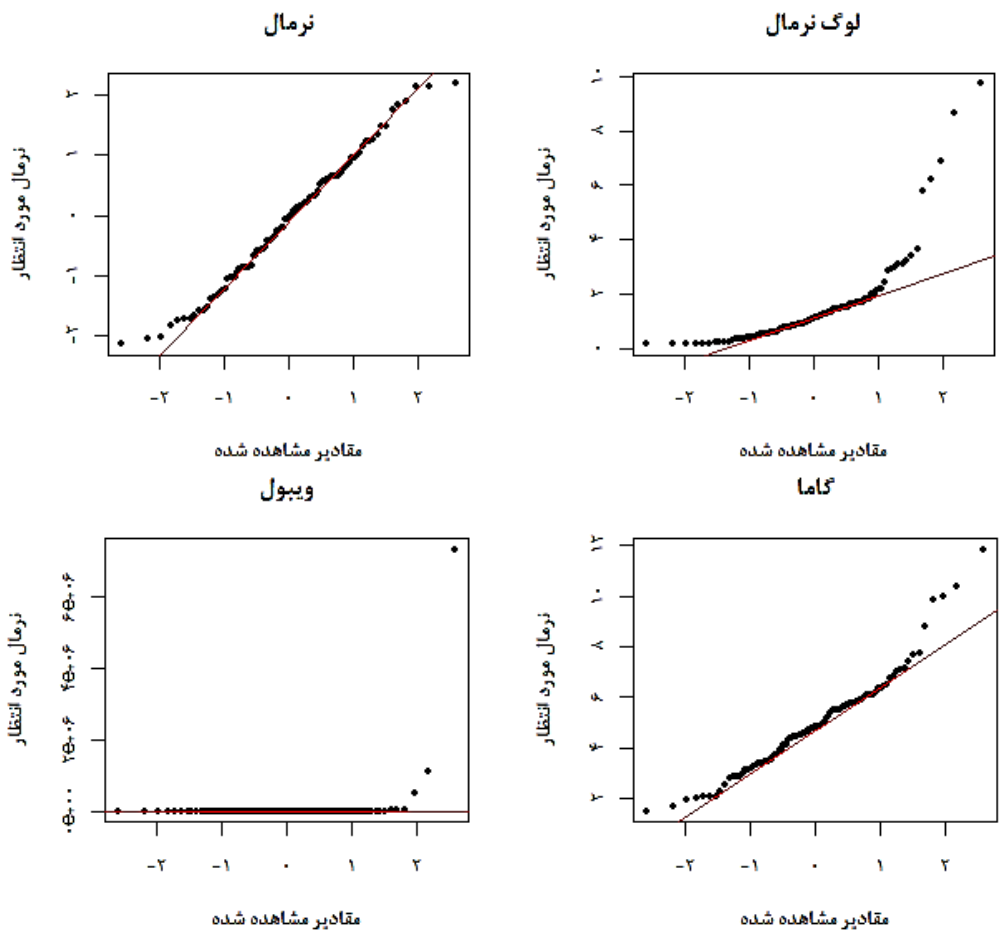
یکی دیگر از روش‌های تصویری به منظور بررسی توزیع نرمال، استفاده از نمودار چندک-چندک یا Q-Q plot است. در این نمودار چندک‌های توزیع تجربی داده‌ها در مقابل چندک‌های نرمال ترسیم می‌شود. هر نقطه از این نمودار به صورت زوج مرتب خواهد بود که مولفه اول مربوط به مقدار چندک توزیع نرمال و مولفه دوم نیز چندک براساس توزیع تجربی داده‌ها است. به این ترتیب رسم این نقاط در مختصات دکارتی، می‌تواند روشی برای تشخیص نرمال بودند توزیع داده‌ها باشد. در صورتی که داده‌ها حول خط ٤٥ درجه قرار بگیرند، داده‌ها نرمال هستند (شکل ٤).



شکل ۲. نمودار هیستوگرام داده‌های چهار توزیع نرمال، لوگ نرمال، گاما و ویبول (N= ٦۰)



شکل ۳. نمودار جعبه‌ای داده‌های چهار توزیع نرمال، لوگ نرمال، گاما و ویبول (N= ٦۰)



شکل ٤. نمودار چندک-چندک داده‌های چهار توزیع نرمال، لوگ نرمال، گاما و ویبول (N= ٦٠)

علاوه بر استفاده از منحنی‌های فوق، می‌توان از آزمون‌های نرمالیتی نیز استفاده نمود. هنگام بررسی نرمال بودن داده‌ها فرض صفر مبتنی بر اینکه توزیع داده‌ها نرمال است را در سطح خطای ۵ درصد مورد آزمون قرار گرفت. بنابراین اگر p.value آزمون بزرگتر مساوی ۰/۰۵ بدست آید، در این صورت دلیلی برای رد فرض صفر مبتنی بر اینکه داده نرمال است، وجود نخواهد داشت. به عبارت دیگر توزیع داده‌ها نرمال خواهد بود. برای آزمون نرمالیته فرض‌های آماری به صورت زیر تنظیم می‌شود:

H_0: توزیع داده‌های مربوط به هر یک از متغیرها نرمال است.

H_1: توزیع داده‌های مربوط به هر یک از متغیرها نرمال نیست.

در پژوهش حاضر از سه آزمون نرمالیتی شاپیرو-ویلک، آندرسون-دارلینگ و لیلی‌فرز (کولموگروف-اسمیرنوف اصلاح شده) استفاده شد.

برازش توزیع داده‌های پوشش گیاهی

با توجه به اینکه آزمون‌های نرمالیتی انجام‌شده صرفاً عدم نرمال بودن داده‌های پوشش گیاهی را تأیید کردند، برای شناسایی الگوی توزیع احتمال حاکم بر داده‌ها، مقادیر درصد پوشش گیاهی در تیپ‌های مختلف با چهار توزیع پیوسته شامل نرمال (به عنوان مدل مبنا)، لوگ‌نرمال، گاما و ویبول برازش داده شد. از میان توزیع‌های پیوسته، صرفاً توزیع‌های دوپارامتری به منظور قابلیت مقایسه یکسان انتخاب گردیدند. به اختصار برخی از

جدول ۲. خلاصه توزیع‌های آماری (اقتباس از ۹)

توزیع	نوع	دامنه	چولگی	مثال
دوجمله‌ای	گسسته	۰ و N	بدون چولگی	تراکم گیاهی
پواسون	گسسته	۰ و ∞	راست تا بدون چولگی	تعداد بذور/گیاهچه در کوادرات
دوجمله‌ای منفی	گسسته	۰ و ∞	راست	تعداد پایه‌های گیاهی در لکه‌ها
هندسی	گسسته	۰ و ∞	راست	دوره زندگی
نرمال	پیوسته	∞ و ∞-	بدون چولگی	pH خاک
گاما	پیوسته	۰ و ∞	راست	غلظت نیتروژن خاک
نمایی	پیوسته	۰ و ∞	راست	زمان بین رخداد بارش
لوگ نرمال	پیوسته	۰ و ∞	راست	وفور گونه‌ای

جدول ۳. انتخاب روش تبدیل داده‌ها براساس میزان چولگی

چولگی	متوسط	زیاد	خیلی زیاد	شدید
مثبت (راست)	ریشه دوم	لگاریتم طبیعی	لگاریتمی در پایه ۱۰	معکوس
منفی (چپ)	معکوس سپس ریشه دوم	معکوس سپس لگاریتم طبیعی	معکوس سپس لگاریتمی در پایه ۱۰	معکوس سپس معکوس

خصوصیات توزیع‌های آماری رایج برای داده‌های گسسته و پیوسته به همراه مثال در پژوهش‌های مرتع در جدول ۲ ارائه شده است.

انتخاب بهترین روش تبدیل داده‌ها

جهت استفاده از آزمون‌های پارامتریک که شرط نرمال بودن داده از پیش فرض‌های اساسی آنها می‌باشد، تبدیل داده‌ها یکی از بهترین راه حل‌ها است. بدین منظور، براساس شکل توزیع (چولگی به راست و چپ) شش روش تبدیل شامل ریشه دوم (جذری)، ریشه سوم، لگاریتمی، لگاریتمی در پایه ۱۰، معکوس و باکس-کاکس بر روی داده‌های پوشش گیاهی اعمال شد (جدول ۳) و پس از تبدیل، مجدداً هیستوگرام داده‌ها ترسیم گردید و نرمال بودن آن‌ها با آزمون شاپیرو-ویلک بررسی شد.

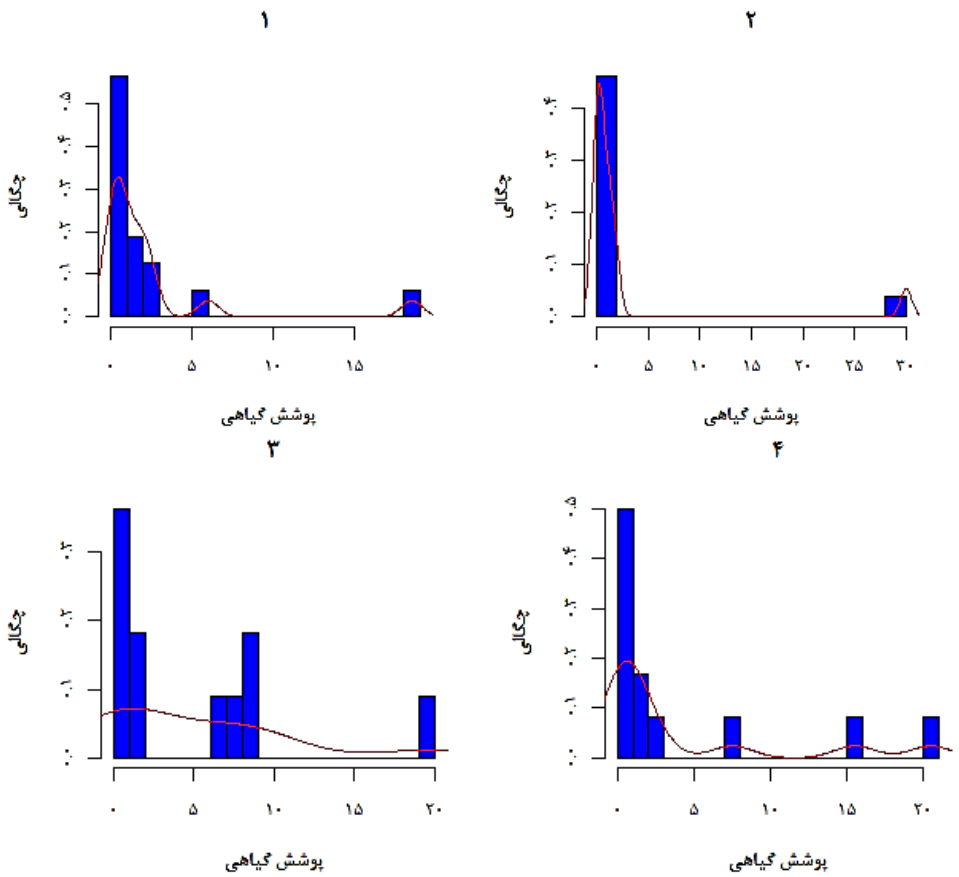
کلیه تحلیل‌های آماری و رسم نمودارها توسط بسته‌های stats، semTools، MASS، nortest، fitdistrplus و graphics در محیط R نسخه ٤.۲.۳ انجام شد (۳۸).

نتایج

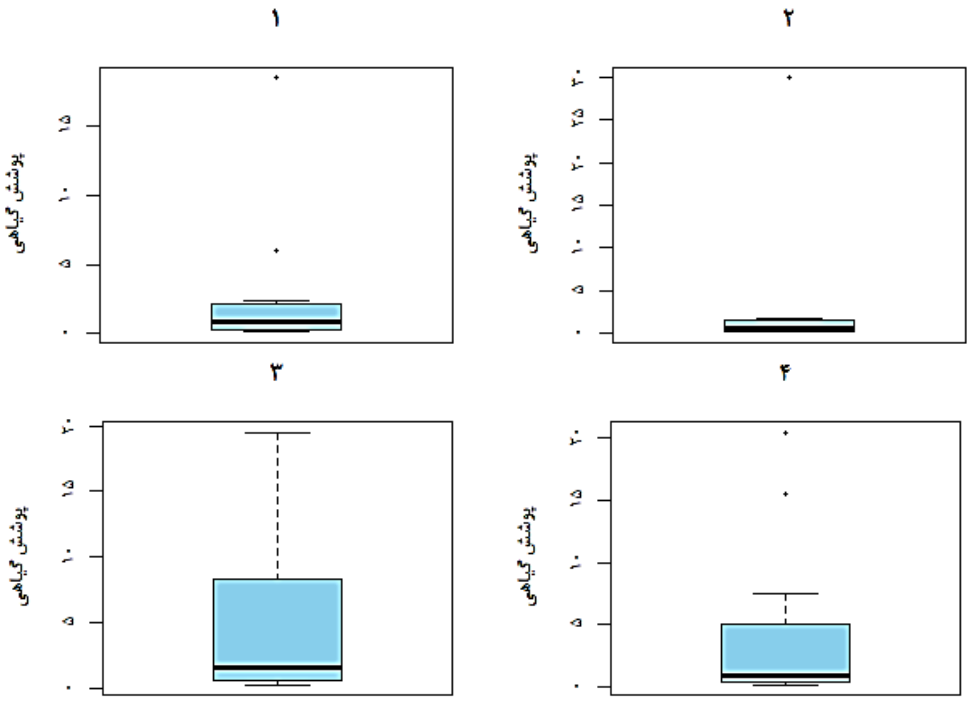
اگرچه کلیه روش‌ها و آزمون‌های آماری روی داده‌های درصد پوشش گیاهی ۱۰ تیپ گیاهی انجام شده است، اما نمودارهای مربوط به چهار تیپ گیاهی منطقه نشان داده می‌شود. نمودار تابع توزیع چگالی چهار تیپ گیاهی (شکل ۵) نشان می‌دهد که در هیچ کدام از تیپ‌ها، داده‌های درصد پوشش گیاهی از توزیع نرمال برخوردار نیستند و با مراجعه به شکل ۱، می‌توان مشاهده کرد که بیشتر تمایل به توزیع لوگ نرمال و ویبول دارند.

نمودار جعبه‌ای داده‌های پوشش گیاهی چهار تیپ گیاهی (شکل ٦) نشان می‌دهد که تمامی تیپ‌ها دارای داده پرت هستند و شکل نمودار جعبه‌ای با الگوی نمودار جعبه‌ای توزیع نرمال (مراجعه شود به شکل ۲) مطابقت ندارد.

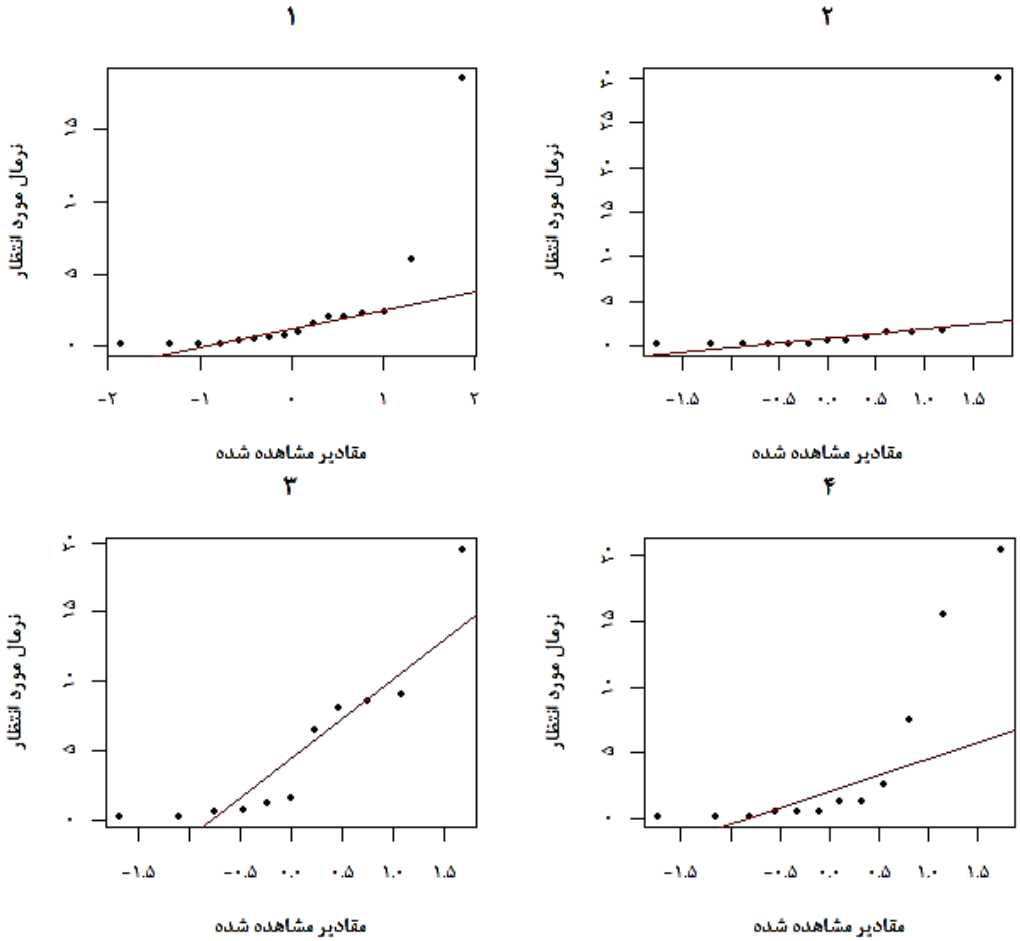
نمودار چندک-چندک داده‌های درصد پوشش گیاهی (شکل ۷) نیز نشان می‌دهد که مقادیر در اطراف خط قرار ندارند و انحراف داده‌های پرت از خط ٤۵ درجه کاملاً مشهود است و این حاکی از رد فرض نرمال بودن داده‌ها است. این نمودارها نیز با



شکل ۵. نمودار تابع توزیع چگالی درصد پوشش گیاهی چهار تیپ گیاهی اول منطقه مورد مطالعه



شکل ٦. نمودار جعبه‌ای درصد پوشش گیاهی چهار تیپ گیاهی اول منطقه مورد مطالعه



شکل ۷. نمودار چندک-چندک درصد پوشش گیاهی چهار تیپ گیاهی اول منطقه مورد مطالعه

نمودارهای چندک-چندک توزیع‌های لوگ نرمال و ویبول مشابه هستند (مراجعه شود به شکل ۳).

آماره‌های توصیفی شامل ضرایب چولگی (گشتاور نوع اول) به همراه اشتباه معیار و نسبت چولگی به اشتباه معیار در جدول ٤ برای ۱۰ تیپ گیاهی محاسبه شده است. همانطور که مشاهده می‌شود ، فقط چولگی تیپ ۵ معنی‌دار نیست($p\geq۰/۰۵$) و حاکی از تقارن داده‌های پوشش گیاهی دارد و بقیه تیپ‌های گیاهی متقارن نبوده ($p\leq۰/۰۵$) و کجی به راست متوسط تا زیاد دارند.

نتیجه آزمون‌های نرمالیتی شاپیرو-ویلک، آندرسون-دارلینگ و لیلی‌فرز (جدول ۵) نشان می‌دهد که بجز تیپ گیاهی ۵، فرض نرمال بودن داده‌های درصد پوشش گیاهی در کلیه تیپ‌های گیاهی رد می‌شود ($p\geq۰/۰۵$).

از آنجایی که نتیجه ضریب چولگی و آزمون‌های نرمالیتی داده‌های پوشش گیاهی تیپ ۵ نشان می‌دهد که فرض نرمال بودن داده‌ها رد نمی‌شود، نمودارهای مربوط به تیپ ۵ در شکل ۱ (پیوست ۱) نشان داده می‌شود، اگر چه نمودار جعبه‌ای و چندک-چندک نشان دهنده نرمال بودن داده‌های این تیپ است، اما نمودار تابع توزیع چگالی گویای انحراف از نرمال بودن است.

از آنجایی که کلیه نمودارها و آزمون‌های فوق نشان می‌دهد که داده‌های درصد پوشش گیاهی از توزیع نرمال برخوردار نیستند، هیستوگرام مقادیر پوشش گیاهی با چهار توزیع نرمال، لوگ نرمال، گاما و ویبول برازش پیدا کرد. به عنوان نمونه، فقط داده-های درصد پوشش گیاهی دو تیپ گیاهی اول در شکل‌های ۲ و ۳ (پیوست ۱) نمایش داده می‌شود. براساس آماره AIC بهترین مدل انتخاب شد. هر چه مقدار AIC کمتر، داده‌های درصد

جدول ٤. ضرایب چولگی (گشتاور نوع اول) به همراه اشتباه معیار (SE) و نسبت چولگی به اشتباه معیار (Z) داده‌های درصد پوشش گیاهی تیپ‌های گیاهی مورد مطالعه

تیپ گیاهی	ضریب چولگی (g_1)	اشتباه معیار چولگی (SE)	z	p.value
١	٣/٣٦	٠/٦١	٥/٥٠	***٠/٠٠٠
٢	٣/٥٧	٠/٦٧	٥/٢٥	***٠/٠٠٠
٣	١/٤٧	٠/٧٤	١/٩٩	*٠/٠٤
٤	١/٨٤	٠/٧١	٢/٦	**٠/٠١
٥	٠/٩٨	٠/٩٢	١/٠٦	٠/٢٩
٦	٢/٦٨	٠/٥٦	٤/٧٧	***٠/٠٠٠
٧	٣/٧٨	٠/٥٨	٦/٥٥	***٠/٠٠٠
٨	٢/٥١	٠/٧٧	٣/٢٤	**٠/٠٠١
٩	١/٦٥	٠/٦٣	٢/٦٢	**٠/٠١
١٠	٢/٥٠	٠/٥٩	٤/٢١	***٠/٠٠٠

*** معنی‌داری در سطح ٠/٠٠١ ($p < 0.001$)، ** معنی‌داری در سطح ٠/٠١ ($p < 0.01$) و * معنی‌داری در سطح ٠/٠٥ ($p < 0.05$) می‌باشد.

جدول ٥. آماره و مقادیر *p* value آزمون‌های آماری بر روی داده‌های درصد پوشش گیاهی تیپ‌های گیاهی مورد مطالعه

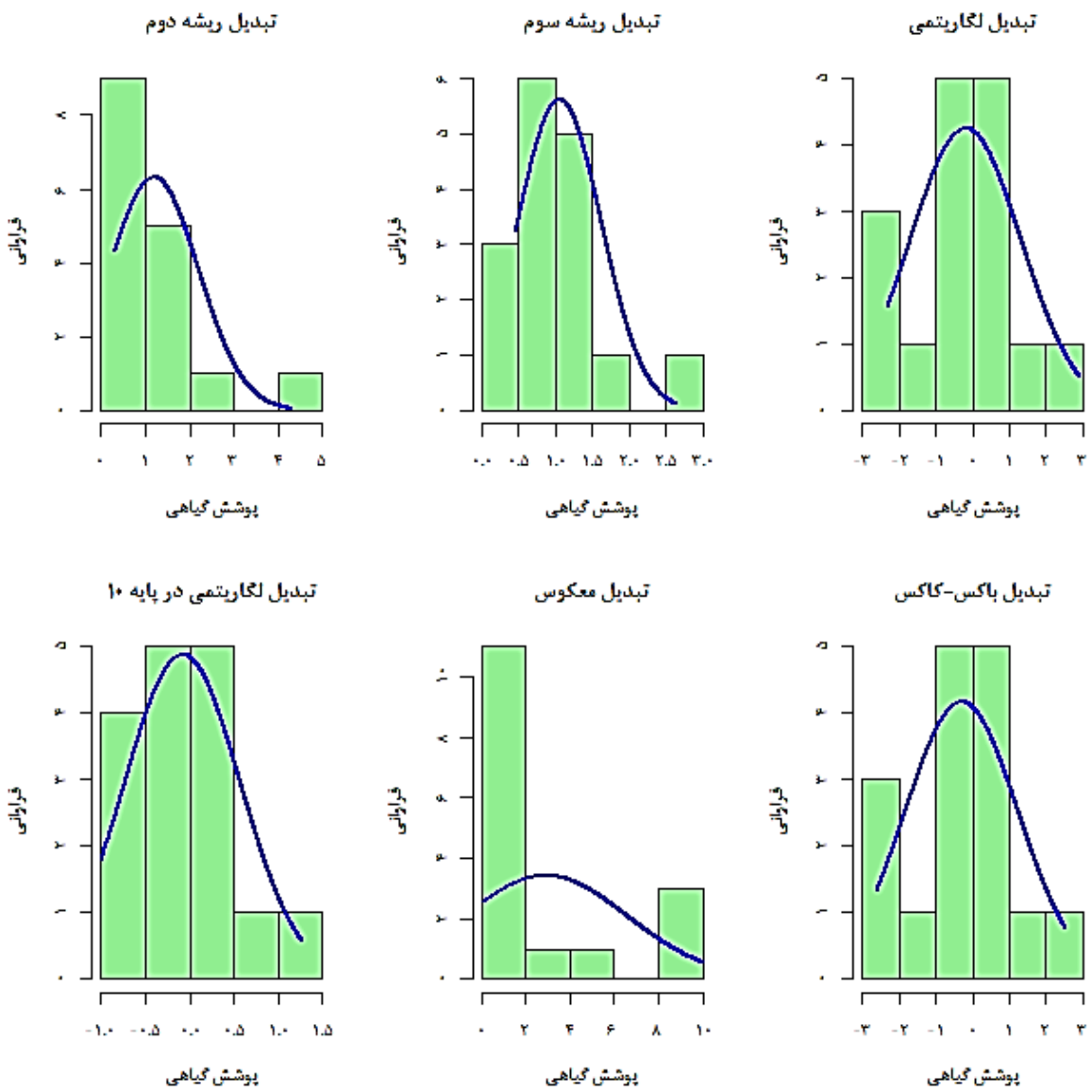
تیپ گیاهی	شاپیرو-ویلک		آندرسون-دارلینگ		لیلی فورز	
	p value	آماره	*p* value	آماره	*p* value	آماره
١	***٠/٠٠	٠/٥٢	***٠/٠٠	٢/٩٥	***٠/٠٠	٠/٣٧
٢	***٠/٠٠	٠/٣٧	***٠/٠٠	٣/٧٤	***٠/٠٠	٠/٤٨
٣	***٠/٠٠	٠/٧٩	*٠/٠٢	٠/٨٨	***٠/٠٠	٠/٢٧
٤	***٠/٠٠	٠/٦٦	***٠/٠٠	١/٨٨	***٠/٠٠	٠/٣٥
٥	٠/٢٨	٠/٨٩	٠/٣٢	٠/٣٦	٠/٥٧	٠/٢٠
٦	***٠/٠٠	٠/٥٢	***٠/٠٠	٤/٠٥	***٠/٠٠	٠/٣٥
٧	***٠/٠٠	٠/٤٨	***٠/٠٠	٣/٣٣	***٠/٠٠	٠/٣٧
٨	***٠/٠٠	٠/٦٥	***٠/٠٠	١/٤٥	**٠/٠١	٠/٣٠
٩	***٠/٠٠	٠/٦٣	***٠/٠٠	٢/٦١	***٠/٠٠	٠/٣٢
١٠	***٠/٠٠	٠/٥٠	***٠/٠٠	٣/٩٩	***٠/٠٠	٠/٤٣

*** معنی‌داری در سطح ٠/٠٠١ ($p < 0.001$)، ** معنی‌داری در سطح ٠/٠١ ($p < 0.01$) و * معنی‌داری در سطح ٠/٠٥ ($p < 0.05$) می‌باشد.

پوشش گیاهی آن تیپ با آن مدل بیشترین برازش را دارند (جدول ٦). در این مطالعه، به منظور ارزیابی برازش چهار توزیع نرمال، لوگ‌نرمال، گاما و ویبول به داده‌های درصد پوشش گیاهی ده تیپ مرتعی، از معیارهای اطلاعات آکائیک (AIC) و اختلاف آکائیک (ΔAIC) استفاده شد. یافته‌ها نشان داد که توزیع لوگ‌نرمال در هشت تیپ شامل تیپ‌های ١، ٢، ٤، ٦، ٧، ٨، ٩ و

جدول ٦. مقادیر حداکثر درستنمایی و آماره اطلاعات آکوئیک (AIC) برازش درصد پوشش گیاهی تیپ‌های گیاهی مورد مطالعه با چهار توزیع نرمال، لوگ نرمال، گاما و ویبول

تیپ گیاهی	نرمال			لوگ نرمال			ویبول			گاما		
	AIC	ΔAIC	Loglikelihood	AIC	ΔAIC	Loglikelihood	AIC	ΔAIC	Loglikelihood	AIC	ΔAIC	Loglikelihood
١	٩٦/٨٢	٤١/٠٤	−٤٦/٤١	٥٥/٧٨	٠/٠٠	−٢٥/٨٩	٥٨/٥٧	٤/٤٦	−٢٧/٢٨	٦٠/٢٤	١/٨٠	−٢٨/١٢
٢	٩٤/٤٧	٦٠/٠٩	−٤٥/٢٣	٣٤/٣٨	٠/٠٠	−١٥/١٩	٣٩/٦٧	٩/٥٧	−١٧/٨٣	٤٣/٩٥	٥/٢٩	−١٩/٩٧
٣	٧٣/٦٢	١٣/٥١	−٣٤/٨١	٦٠/١٨	٠/٠٧	−٢٨/٠٩	٦٠/١١	٠/٠٤	−٢٨/٠٦	٦٠/١٥	٠/٠٠	−٢٨/٠٧
٤	٨٣/٢٠	٣٠/٨٨	−٣٩/٦٠	٥٢/٣٢	٠/٠٠	−٢٤/١٦	٥٤/١١	٣/١٩	−٢٥/٠٥	٥٥/٥١	١/٧٩	−٢٥/٦٠
٥	٥١/٤٢	٥/٨٨	−٢٣/٧١	٤٦/٨٩	١/٣٥	−٢١/٤٤	٤٥/٥٤	٠/٢١	−٢٠/٩٢	٤٥/٧٥	٠/٠٠	−٢٠/٨٧
٦	١٢٠/٦٥	٦٣/٣٠	−٥٨/٣٢	٥٧/٣٥	٠/٠٠	−٢٦/٦٧	٦٤/٨٨	١١/٣٠	−٣٠/٤٤	٦٨/٦٥	٧/٥٣	−٣٢/٣٢
٧	٩٧/١٨	٤٠/٩١	−٤٦/٥٩	٥٦/٢٧	٠/٠٠	−٢٦/١٣	٦٠/٥٧	٥/٥٠	−٢٨/٢٨	٦١/٧٧	٤/٣٠	−٢٨/٨٩
٨	٧٣/٢٥	١٨/٠٧	−٣٤/٦٢	٥٥/١٨	٠/٠٠	−٢٥/٥٩	٥٦/٢١	١/٥١	−٢٦/١٠	٥٦/٦٩	١/٠٣	−٢٦/٣٤
٩	٩٥/٧٩	٤٠/١١	−٤٥/٨٩	٥٥/٦٨	٠/٠٠	−٢٥/٨٤	٥٩/٤٧	٥/٨٤	−٢٧/٧٣	٦١/٥٢	٣/٧٩	−٢٨/٦١
١٠	٩٦/٢٨	٥٥/٣٢	−٤٦/١٤	٤٠/٩٦	٠/٠٠	−١٨/٤٨	٤٨/٩٠	١١/٣٤	−٢٢/٤٥	٥٢/٣٠	٧/٩٤	−٢٤/١٥



شکل ۸. هیستوگرام داده‌های تبدیل شده درصد پوشش گیاهی تیپ اول به روش‌های مختلف

۱۰ دارای کمترین مقدار AIC و در نتیجه بهترین برازش بود، در حالی که توزیع ویبول در دو تیپ ۳ و ۵ بهترین برازش را نشان داد. توزیع‌های نرمال و گاما در هیچ تیپی به عنوان بهترین مدل انتخاب نشدند.

حال که نتایج، حاکی از رد فرض نرمال بودن داده‌های درصد پوشش گیاهی است، جهت استفاده از آزمون‌های آماری پارامتریک، روش‌های مختلف تبدیل داده‌ها روی داده‌های درصد پوشش گیاهی اعمال شد. به عنوان نمونه برای تیپ گیاهی اول، نتایج نشان می‌دهد که انجام تبدیل لگاریتم طبیعی، لگاریتم در پایه ۱۰ و باکس-کاکس منجر به نرمال شدن داده‌ها شد (شکل ۸).

برای ۱۰ تیپ گیاهی، مقادیر *p* value آزمون نرمالیتی شاپیرو-ویلک بعد از تبدیل داده‌ها در جدول ۷ ارائه شده است. همانطور که مشاهده می‌شود، از ۱۰ تیپ گیاهی، فرض نرمال بودن داده‌های درصد پوشش گیاهی در ۵ تیپ گیاهی پس از تبدیل داده‌ها رد نشد ($p \geq ۰/۰۵$). فرض نرمال بودن داده‌های پوشش

جدول ۷. مقادیر *p* value آزمون شاپیرو-ویلک بر روی داده‌های درصد پوشش گیاهی تبدیل شده به روش‌های مختلف مربوط به تیپ‌های گیاهی مورد مطالعه

تیپ گیاهی	ریشه دوم	ریشه سوم	لگاریتمی	لگاریتمی در پایه ۱۰	معکوس	باکس-کاکس
۱	***۰/۰۰	*۰/۰۲	۰/۵۳	۰/۵۳	***۰/۰۰	۰/٤٤
۲	***۰/۰۰	***۰/۰۰	*۰/۰۲	*۰/۰۲	***۰/۰۰	*۰/۰٤
۳	۰/۱۱	۰/۱۸	۰/۲۳	۰/۲۳	***۰/۰۰	۰/۲٤
٤	***۰/۰۰	*۰/۰۳	۰/۲٤	۰/۲٤	***۰/۰۰	۰/۲٤
۵	۰/۸۵	۰/۹۳	۰/٦۷	۰/٦۷	***۰/۰۰	۰/۹۱
٦	***۰/۰۰	***۰/۰۰	***۰/۰۰	***۰/۰۰	***۰/۰۰	***۰/۰۰
۷	***۰/۰۰	***۰/۰۰	*۰/۰٤	*۰/۰٤	***۰/۰۰	*۰/۰٤
۸	۰/۰۷	۰/٤۱	۰/۹۹	۰/۹۹	***۰/۰۰	۰/۹۹
۹	***۰/۰۰	***۰/۰۰	***۰/۰۰	***۰/۰۰	***۰/۰۰	***۰/۰۰
۱۰	***۰/۰۰	***۰/۰۰	***۰/۰۰	***۰/۰۰	*۰/۰۱	*۰/۰۱

*** معنی‌داری در سطح ۰/۰۰۱ ($p < 0.001$)، ** معنی‌داری در سطح ۰/۰۱ ($p < 0.01$) و * معنی‌داری در سطح ۰/۰۵ ($p < 0.05$) می‌باشد.

گیاهی برخی از تیپ‌ها با اعمال هیچکدام از تبدیل‌ها نیز رد شد ($p \leq ۰/۰۵$). نتیجه تبدیل لگاریتمی، لگاریتم در پایه ۱۰ و باکس-کاکس مشابه هستند. تبدیل معکوس اصلاً برای این داده‌ها مناسب نبود.

بحث و نتیجه‌گیری

پژوهش حاضر با استفاده از رویکردی تلفیقی شامل آزمون‌های آماری، شاخص‌های توصیفی، نمودارهای گرافیکی و برازش توزیع‌ها با معیار AIC نشان داد که داده‌های درصد پوشش گیاهی در ۹۰ درصد تیپ‌های مرتعی منطقه حفاظت‌شده درمیان-سربیشه از توزیع نرمال پیروی نمی‌کنند و عمدتاً دارای چولگی راست هستند. توزیع‌های لوگ‌نرمال (۸۰ درصد تیپ‌ها) و ویبول (۲۰ درصد تیپ‌ها) بهترین برازش را داشتند. این یافته با نظر Popovic و همکاران (۲۰۲۲) و Krebs (۲۰۱٤) مبنی بر عدم نرمال بودن اغلب داده‌های بوم‌شناختی همخوانی دارد (۳۷ و ۲٤). با این حال، در پژوهش‌هایی که اثر تیمارهای مختلف بر پوشش گیاهی بررسی می‌شود، بر اساس قضیه حد مرکزی، توزیع میانگین نمونه‌ای تقریباً نرمال خواهد بود و نمی‌توان با قطعیت گفت که داده‌ها نرمال نیستند.

استفاده از روش‌های متعدد سنجش نرمال بودن به دلیل محدودیت‌های هر روش ضروری بود؛ آزمون‌های نرمالیتی در نمونه‌های کوچک توان پایینی دارند (٤٤) و در نمونه‌های بزرگ ممکن است انحرافات جزئی را معنی‌دار نشان دهند (٤۸). بنابراین ترکیب ارزیابی گرافیکی و عددی به نتیجه‌گیری صحیح‌تر کمک می‌کند. نتایج نشان داد که تبدیل‌های لگاریتمی و باکس-کاکس موفق‌ترین روش‌ها در نرمال‌سازی داده‌ها بودند، اما تنها در نیمی از تیپ‌ها مؤثر واقع شدند که با ادعای Stohlgren (۲۰۰۷) مبنی بر اینکه تبدیل داده‌ها لزوماً باعث نرمال شدن نمی‌شود، مطابقت دارد؛ حتی در برخی شرایط، تبدیل داده‌ها ممکن است اختلاف بین گروه‌ها را کوچکتر کند و تشخیص تفاوت معنی‌دار دشوارتر شود (۵۳). Zar (۲۰۱۰) تبدیل log10(x+1) را برای درصد پوشش توصیه می‌کند و تبدیل باکس-کاکس به دلیل انعطاف‌پذیری بالا در منابع متعددی مورد تأکید قرار گرفته است (۱۷، ۱۹، ۲۵ و ۲٦). در داده‌های جامعه‌شناسی گیاهی، روش‌های استانداردسازی (مرکز کردن و

جدول ۸. اولویت‌بندی روش‌های تبدیل داده برای نرمال‌سازی درصد پوشش گیاهی در مناطق با اقلیم خشک

اولویت	روش تبدیل	میزان موفقیت در این پژوهش	توصیه کاربردی
۱	باکس-کاکس	بیشترین موفقیت	مناسب برای درجات مختلف چولگی، برآورد بهینهλ
۲	لگاریتمی (Ln یا Log10)	موفقیت بالا (مشابه باکس-کاکس)	مناسب برای داده‌های با توزیع لوگ‌نرمال، استفاده از Ln(x+1) برای داده‌های حاوی صفر
۳	ریشه سوم	موفقیت متوسط	مناسب برای چولگی متوسط و داده‌های حاوی صفر
٤	ریشه دوم	موفقیت محدود	مناسب برای داده‌های شمارشی مثل تراکم (نه درصد پوشش)
نامناسب	معکوس	موفقیت صفر	هرگز توصیه نمی‌شود.

نرمال‌سازی) که در نرم‌افزارهایی مثل PC-ORD و CANOCO (٤١) و بسته‌های R (۳۵) پیش‌فرض هستند، موقعیت و پراکنش داده‌ها را تغییر می‌دهند اما شکل توزیع را تغییر نمی‌دهند (٦) و نباید با تبدیل داده‌ها اشتباه شوند.

در انتخاب روش تبدیل، علاوه بر معیار چولگی، قوانین کلی مانند استفاده از ریشه دوم برای داده‌های ۳۰-۰ درصد و ۷۰- ۱۰۰ درصد، و تبدیل آرک سین برای سایر محدوده‌ها پیشنهاد می‌شود (١٦). همچنین هدف پژوهش تعیین‌کننده است؛ برای نرمال‌سازی و همگن‌سازی واریانس، تبدیل‌های ریشه‌ای و لگاریتمی و برای خطی‌سازی روابط نمایی، تبدیل لگاریتمی مناسب است (١١). در نهایت بر اساس یافته‌های پژوهش حاضر، اولویت‌بندی روش‌های تبدیل داده برای مناطق با اقلیم خشک به شرح جدول ۸ است.

اگرچه نرمالیتی یکی از پیش‌شرط‌های اصلی تحلیل‌های آماری است، برخی آزمون‌ها مانند ANOVA در برابر انحرافات جزئی تا متوسط مقاوم هستند (٤٦) و انحراف جزیی تاثیر جدی بر عملکرد آزمون نمی‌گذارد (۳). اما اگر داده‌ها نرمال نباشند، میانگین نماینده داده‌ها نخواهد بود و مقایسه میانگین‌ها پس از ANOVA بی‌اعتبار خواهد بود (۳۰). در تحلیل‌های چندمتغیره، روش‌های خوشه‌بندی و رسته‌بندی مانند DCA و NMDS نیاز به نرمال بودن ندارند (٢٦)، اما PCA به عنوان یکی از پرکاربردترین روش‌ها، به تقارن داده‌ها حساس است (٤) و انحراف زیاد از تقارن عملکرد آن را تحت تأثیر قرار می‌دهد (۱۸). از این رو در حالی که PCA می‌تواند بسیار مفید باشد اما برای جلوگیری از نتایج گمراه کننده، مراقبت زیادی لازم است (۷). همچنین روش CCA از رویکرد ANOVA/رگرسیون استفاده می‌کند (٤٣) و نرمال بودن در آن اهمیت دارد. از بین روش‌های تحلیل عاملی، روش حداکثر درست‌نمایی (ML) به نرمالیتی چندمتغیره نیاز دارد (۱۲) و نرمالیتی چندمتغیره برای تحلیل همبستگی متعارفی (CCA) در مطالعات خاک و گیاه ضروری است (۲۸).

برای مقابله با عدم نرمال بودن، روش‌های مقاوم مانند Robust PCA، Robust PLS (۱۳)، رسته‌بندی گوسی براساس تابع مفصل (۳۷) و رسته‌بندی تطبیقی گوسی (CGO) به جای CCA پیشنهاد می‌شوند (٤٩). همچنین مدل‌های رگرسیون بتا و مدل‌های صفر آماسیده (Zero-inflated) برای داده‌های درصدی و داده‌های دارای صفر فراوان گزینه‌های مناسبی هستند.

در پاسخ به پرسش اصلی پژوهش، در مناطق خشک با پوشش ناهمگن، الگوی پراکنش کپه‌ای، نمونه‌برداری تصادفی و تخمین چشمی، انتظار نرمال نبودن داده‌ها وجود داشت. دو محدودیت اصلی پژوهش شامل تعداد ٦٠ پلات در هر تیپ و اقلیم نیمه‌خشک منطقه است که تعمیم نتایج را محدود می‌کند. در نهایت، انجام پژوهش‌های تطبیقی در اقلیم‌های مختلف، بررسی اثر اندازه پلات، استفاده از رویکردهای بیزی و طراحی

مطالعات شبیه‌سازی برای تدوین راهنمای عملی دقیق‌تر پیشنهاد می‌شود. پیام اصلی آنکه محققان علوم مرتع پیش از استفاده از آزمون‌های پارامتریک، نرمال بودن داده‌ها را با رویکردی ترکیبی بررسی کرده و در صورت عدم تایید، از روش‌های آماری مقاوم یا مدل‌های جایگزین استفاده نمایند.

تشکر و قدردانی

بدینوسیله از معاونت محترم پژوهشی، فناوری و نوآوری دانشگاه بیرجند و کارشناسان محترم اداره کل حفاظت محیط زیست استان خراسان جنوبی به دلیل همکاری در دریافت اطلاعات و راهنمایی‌های لازم در انجام این پژوهش، تشکر و قدردانی می‌گردد.

منابع

1. Amiri, F., Arzani, H., Farahpour, M., Chaeichi, M.R. and Khajeh Aldin, S.J.A.D., 2009. Efficiency of MPSIAC and EPM Models for Assessment Soil Erosion in Range Suitability. *Journal of Rangeland Science*, 3(1): 138-154. (In Persian)
2. Arzani, H. and Abedi, M., 2013. Rangeland assessment survey and monitoring. University of Tehran Press, Tehran. 305p. (In Persian)
3. Bhattacharya, B. and Habtzghi, D., 2002. Median of the p-value under the alternative hypothesis. *The American Statistician*, 56(3): 202-206.
4. Bialik, O.M., Jarochowska, E. and Grossowicz, M., 2021. Ordination analyses in sedimentology, geochemistry and paleoenvironment - current trends and recommendations. EGU General Assembly 2021, Online (Virtual), pp. 19-30, EGU21-3363.
5. Bihamta, M.R. and Zare Chahouki, M.A., 2015. Principles of statistics for the natural resources science. 4th Edition. University of Tehran Press, Tehran. 302p. (In Persian)
6. Birks, H.J.B., Lotter, A.F., Juggins, S. and Smol, J.P. (Eds.), 2012. Tracking environmental change using lake sediments: developments in paleoenvironmental research. Springer Netherlands, Dordrecht.
7. Björklund, M., 2019. Be careful with your principal components. *Evolution: International Journal of Organic Evolution*, 73(10): 2151-2158.
8. Blanca, M.J., Alarcón, R., Arnau, J., Bono, R. and Bendayan, R., 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema*, 29(4): 552-557.
9. Bolker, B., 2008. Probability and stochastic distributions for ecological modeling. pp. 103-146. *In: Ecological Models and Data in R*. Princeton University Press, Princeton.
10. Bonham, C.D., 2013. Measurements for terrestrial vegetation. John Wiley and Sons, New York.
11. Borcard, D., Gillet, F. and Legendre, P., 2018. Numerical ecology with R. 2nd Edition. Springer International Publishing, Switzerland.
12. Brown, T.A., 2006. Confirmatory factor analysis for applied research. Guilford Press, New York.
13. Brown, S., Tauler, R. and Walczak, B., 2020. Comprehensive chemometrics: chemical and biochemical data analysis. 2nd Edition. Elsevier, Amsterdam.
14. Callaghan, C.T., Borda-de-Água, L., Van Klink, R., Rozzi, R. and Pereira, H.M., 2023. Unveiling global species abundance distributions. *Nature ecology & evolution*, *7*(10): 1600-1609.
15. Ghasemi, A. and Zahediasl, S., 2012. Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2): 486-489.
16. Gomez, K.A. and Gomez, A.A., 1984. Statistical procedures for agricultural research. 2nd Edition. John Wiley and Sons, New York. 680p.
17. Greenacre, M.J. and Primicerio, R., 2013. Multivariate analysis of ecological data. Fundación BBVA, Bilbao. 336p.
18. Hammer, Ø. and Harper, D.A.T., 2007. Paleontological data analysis. Blackwell Publishing Ltd, Hoboken, New Jersey. 376p.
19. Henderson, P.A. and Southwood, T.R.E., 2016. Ecological methods. 4th Edition. Blackwell Science, USA. 656p.
20. Jafari, M. and Rostampour, M., 2019. Soil - Plant relationships: ecology, statistics and analysis. Vol. 1. University of Tehran Press, Tehran. (In Persian)
21. Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B., Wright, I.J. and Cornelissen, J.H.C., 2011. TRY–a global database of plant traits. *Global change biology*, *17*(9): 2905-2935.

22. Kent, M., 2012. Vegetation Description and Data Analysis: A practical approach. 2nd Edition. Wiley-Blackwell, Chichester, UK. 403p.
23. Korkmaz, S., Goksuluk, D. and Zararsiz, G., 2014. MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2): 151-162.
24. Krebs, C.J., 2014. Ecological methodology. 3rd Edition. Addison-Wesley Educational Publishers, Inc., Boston, USA. 620p.
25. Legendre, P. and Borcard, D., 2018. Box-Cox-chord transformations for community composition data prior to beta diversity analysis. *Ecography*, 41(11): 1820–1824.
26. Legendre, P. and Legendre, L., 2012. Numerical ecology. 3rd Edition. Elsevier, Amsterdam. 1006p.
27. Magurran, A.E. and McGill, B.J., 2011. Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford. 368p.
28. Menzel, U., 2022. CCP: Significance tests for Canonical Correlation Analysis (CCA). R package version 1.2. Available online at: https://CRAN.R-project.org/package=CCP. Accessed 13 July 2026.
29. Mesdaghi, M., 2014. Quantitative ecology. Mashhad Academic Jihad Publications, Mashhad. (In Persian)
30. Mishra, P., Pandey, C.M., Singh, U., Gupta, A., Sahu, C. and Keshri, A., 2019. Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1): 67–72.
31. Moghaddam, M.R., 2008. Statistical and Descriptive Ecology of Vegetation. Tehran University Press, Tehran. (In Persian).
32. Moghaddam, M.R., 2014. Range and rang management. 7th Edition. University of Tehran Press, Tehran. (In Persian).
33. Molaeinasab, A., Bashari, H. and Tarkesh Esfahani, M., 2018. Comparing rangeland functions using landscape function analysis and rangeland health (Case study: Meidanak rangelands, Fereidoonshar). *Journal of Rangeland*, 12(2): 242-252. (In Persian)
34. Motamedi, J., 2023. Suitability assessment of mountain rangelands in the Marghevar region of Urmia for dual-purpose grazing and medicinal plant exploitation. *Journal of Rangeland*, 17(2): 312-333. (In Persian)
35. Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. and Oksanen, M.J., 2019. vegan: Community ecology package. Available online at: https://cran.r-project.org/web/packages/vegan/index.html. Accessed 13 July 2026.
36. Palmer, M.W., 2019. Gradient analysis of ecological communities (Ordination). pp. 241-274. *In: Gelfand, A., Fuentes, M., Hoeting, P. and Smith, R.L. (Eds.), Handbook of Environmental and Ecological Statistics*. CRC Press, Boca Raton.
37. Popovic, G.C., Hui, F.K.C. and Warton, D.I., 2022. Fast model-based ordination with copulas. *Methods in Ecology and Evolution*, 13: 194–202.
38. R Core Team, 2025. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Available online at: https://www.R-project.org/. Accessed 13 July 2026.
39. Rostampour, M., Mashgani, M. and Akbari, H., 2019. A study of floristic, functional, and relative diversity of plant families in Darmian and Sarbisheh protected area. *Taxonomy and Biosystematics*, 11(39): 75-96. (In Persian)
40. Rostampour, M. and Sabzi, R., 2022. Effect of slope gradient on vegetation, species composition and production of medicinal plants (Case study: Room Rangelands, Qaen). *Journal of Rangeland*, 16(1): 312-330. (In Persian)
41. Šmilauer, P. and Lepš, J., 2014. Multivariate Analysis of Ecological Data Using CANOCO 5. 2nd Edition. Cambridge University Press, Cambridge, UK. 269p.
42. Stohlgren, T.J., 2007. Measuring plant diversity: lessons from the field. Oxford University Press, Oxford. 410p.
43. ter Braak, C. and Šmilauer, P., 2015. Topics in constrained and unconstrained ordination. *Plant Ecology*, 216(5): 683-696.
44. Tsagris, M. and Pandis, N., 2021. Normality test: is it really necessary? *American Journal of Orthodontics and Dentofacial Orthopedics*, 159(4): 548-549.
45. Ulrich, W., Matthews, T.J., Biurrun, I. and et al., 2022. Environmental drivers and spatial scaling of species abundance distributions in Palaearctic grassland vegetation. *Ecology*, 103(8): e3725.
46. Wilcox, R., 2012. Introduction to robust estimation and hypothesis testing. 3rd Edition. Academic Press, Boston. 690p.
47. Wildi, O., 2017. Data Analysis in vegetation ecology. 3rd Edition. CABI, Wallingford. 306p.
48. Yang, S. and Berdine, G., 2021. Normality tests. *The Southwest Respiratory and Critical Care Chronicles*, 9(37): 87-90.
49. Yee, T.W., 2004. A new technique for maximum-likelihood canonical Gaussian ordination. *Ecological Monographs*, 74(4): 685-701.

50. Zar, J.H., 2010. Biostatistical analysis. Pearson, London, UK. 960p.
51. Chahouki, M.A.Z., 2013. Classification and ordination methods as a tool for analyzing of plant communities. In *Multivariate analysis in management, engineering and the sciences*. IntechOpen.
52. Zuur, A.F., Ieno, E.N. and Smith, G.M., 2007. Analyzing ecological data. Statistics for biology and health. Springer New York, New York. 698p.
53. Zuur, A.F., Ieno, E.N. and Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1): 3-14.

Assessment of Statistical Distribution and Data Transformation Methods for Vegetation Cover Percentage in the Rangelands of the Darmian-Sarbisheh Protected Area

Moslem Rostampour

(Received: February 04-2026; Accepted: July 08-2026)

Abstract

This study examined the statistical distribution of vegetation cover percentages across ten rangeland types within the Darmian-Sarbisheh Protected Area and evaluated the efficacy of various data transformation techniques. Vegetation cover was quantified in 600 plots of 4 m² each. Normality assessments employed skewness coefficients, graphical analyses, and formal tests-including Shapiro-Wilk, Anderson-Darling, and Lilliefors tests. Results revealed that nine of the ten vegetation types exhibited significant right-skewness and non-normal distributions, with only one type conforming to normality. Distribution fitting using the Akaike Information Criterion indicated that the normal distribution was inadequate for approximately 90% of the data, whereas log-normal and Weibull distributions provided better fits-accounting for 80% and 20% of the types, respectively. Among six transformation methods (square root, cube root, natural and base-10 logarithms, inverse, and Box-Cox), only logarithmic and Box-Cox transformations effectively normalized 50% of the datasets; the inverse transformation was entirely ineffective. Given the pervasive non-normality in semi-arid rangeland vegetation cover data and the limited success of transformations, the application of robust statistical techniques-such as Robust Principal Component Analysis-is recommended for ecological and rangeland management studies, ensuring more reliable analytical outcomes.

Keywords: Multivariate Statistics, Vegetation, Log-Normal Distribution, Goodness of Fit.

Associate Professor, Department of Rangeland and Watershed Management, Faculty of Natural Resources and Environment, University of Birjand, Birjand, Iran

* Corresponding Author, Email: rostampour@birjand.ac.ir