ارزیابی روش‌های گروه‌بندی ژنوتیپ‌های کلزا با استفاده از تجزیه تابع تشخیص خطی فیشر

بابک ربیعی* و مهدی رحیمی[۱]

(تاریخ دریافت: ۱۳۸۷/۶/۱۷ ؛ تاریخ پذیرش: ۱۳۸۷/۱۲/۲۶)

چکیده

تجزیه تابع تشخیص یکی از روش‌های تجزیه آماری چند متغیره است که از آن می‌توان برای آزمون صحت نتایج حاصل از تجزیه خوشه‌ای استفاده نمود. در این مطالعه، صحت گروه‌بندی روش‌های مختلف تجزیه خوشه‌ای بر پایه روش‌های مختلف استاندارد کردن داده‌ها و معیارهای متفاوت فاصله با تجزیه تابع تشخیص مورد ارزیابی قرار گرفتند. هم‌چنین برای تأیید نتایج از T^2 هتلینگ، پلات CCC و تجزیه واریانس چند متغیره استفاده گردید. بدین منظور، ۸ ژنوتیپ کلزا در قالب طرح بلوک‌های کامل تصادفی با سه تکرار در مؤسسه تحقیقات برنج کشور (رشت) در سال ۸۵-۱۳۸۴ کشت شدند و ۱۴ صفت در آنها مورد ارزیابی قرار گرفت. تجزیه واریانس طرح بلوکی اختلاف معنی‌داری را بین ژنوتیپ‌ها از نظر کلیه صفات مورد مطالعه نشان داد. مقایسه میانگین بین ژنوتیپ‌ها نیز نشان داد که ژنوتیپ Hyola401 از نظر عملکرد دانه و بسیاری از صفات بررسی شده برتر از سایر ژنوتیپ‌ها بود. براورد ضریب تغییرات فنوتیپی و ژنوتیپی نشان داد که اکثر صفات دارای تنوع زیادی در جمعیت می‌باشند. تجزیه تابع تشخیص نشان داد که معیار فاصله اقلیدسی بهتر از سایر معیارهای فاصله بود و گروه‌بندی مطلوبی بر اساس آن به‌دست آمد. هم‌چنین تمام روش‌های استاندارد کردن داده‌ها گروه‌بندی مشابهی به‌وجود آوردند و بهتر از استاندارد نکردن داده‌ها بودند. بر اساس ارزیابی دندروگرام‌های روش‌های مختلف تجزیه خوشه‌ای مشخص شد که روش‌های متوسط فاصله بین گروه‌ها (UPGMA)، دورترین همسایه‌ها و حداقل واریانس "وارد" بهتر از سایر روش‌ها بودند و ژنوتیپ‌ها را در سه گروه دسته‌بندی کردند. تجزیه تابع تشخیص خطی فیشر نشان داد که روش‌های UPGMA و حداقل واریانس "وارد" با انجام صحت گروه‌بندی در حدود ۸۷/۵ درصد، مناسب‌تر از سایر روش‌های تجزیه خوشه‌ای بودند، با این حال تجزیه تشخیص ژنوتیپ‌ها را در دو گروه قرار داد. آزمون‌های T^2 هتلینگ، پلات CCC و تجزیه واریانس چند متغیره نیز نتایج حاصل از تجزیه تشخیص را مورد تأیید قرار دادند. به این ترتیب، به‌نظر می‌رسد که استفاده از معیار فاصله اقلیدسی بر اساس داده‌های استاندارد شده و انجام تجزیه خوشه‌ای با روش‌های حداقل واریانس "وارد" و یا UPGMA گروه‌بندی بهتری از ژنوتیپ‌ها ارائه دهد، اما توصیه می‌شود برای تأیید نتایج و تعیین گروه‌های واقعی از تجزیه تابع تشخیص استفاده گردد.

واژه‌های کلیدی: استاندارد کردن داده‌ها، تجزیه تابع تشخیص خطی فیشر، تجزیه خوشه‌ای، کلزا

مقدمه

برای آن که به‌نژادگر بتواند حداکثر بهره‌برداری را از پدیده هتروزیس به‌عمل آورد، ابتدا لازم است میزان تنوع موجود در بین ژنوتیپ‌های مورد مطالعه را ارزیابی نماید وسپس با دورگ‌گیری بین ژنوتیپ‌هایی که از نظر تنوع، تفاوت عمده‌ای با یک‌دیگر دارند به هیبریدهای پر محصول و با صفات مطلوب دست یابد. تجزیه خوشه‌ای یکی از روش‌های آماری چند متغیره است که برای تعیین تنوع بین جوامع مختلف گیاهی و

۱. به‌ترتیب استادیار و دانشجوی سابق کارشناسی ارشد زراعت و اصلاح نباتات، دانشکده علوم کشاورزی، دانشگاه گیلان

* : مسئول مکاتبات، پست الکترونیکی: rabiei@guilan.ac.ir

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

جانوری و دسته‌بندی آنها به گروه‌های مختلف بر اساس فاصله یا تشابه ژنتیکی مورد استفاده قرار می‌گیرد (۱۲). این روش حداقل در دو مورد می‌تواند به به‌نژادگر کمک نماید: یکی پیدا کردن گروه‌های واقعی افراد بر اساس تشابه ژنتیکی بین آنها و دیگر کاهش داده‌ها و انتخاب افراد محدودی از هر گروه یا دسته (۹).

اما مشکل عمده و اساسی استفاده از روش تجزیه خوشه‌ای در گروه‌بندی افراد یا ژنوتیپ‌ها این است که روش‌های بسیار متفاوتی برای انجام این نوع تجزیه توسط محققین مختلف پیشنهاد شده است که در بسیاری از موارد نتایج متفاوتی نیز ارائه می‌دهند. به این ترتیب تشخیص صحت و سقم نتایج حاصل از روش‌های مختلف و گروه‌بندی‌های حاصل به‌وسیله محقق، بسیار مشکل و گمراه کننده خواهد بود. یکی از روش‌هایی که از آن برای تشخیص صحت گروه‌بندی به‌دست آمده از تجزیه‌ی خوشه‌ای می‌توان استفاده نمود، تجزیه تابع تشخیص است که برای این منظور می‌توان از روش‌های کنارگذاری (Hold out) و u اعتباری (Cross-Validation U-method) استفاده نمود (۸ و ۹). کارایی معیار تشخیص را هم می‌توان به‌وسیله تخمین احتمالات گروه‌بندی نادرست از مشاهدات جدید ارزیابی کرد (۶). مندز و همکاران (۱۰) با استفاده از تجزیه تشخیص بر روی ۹ گروهی که از تجزیه خوشه‌ای با روش متوسط فاصله بین و درون کلاسترها با معیار فاصله‌ی پیرسون به‌دست آمده بود، نشان دادند که از ۹ گروه اولیه، فقط دو گروه ۱۰۰ درصد صحیح گروه‌بندی شده و ۵ گروه تنها در حدود ۶۰ درصد صحیح گروه‌بندی شده بودند. موردا و همکاران (۱۱) نیز با تجزیه تابع تشخیص نشان دادند که ۹۴/۴ درصد از گروه‌بندی حاصل از تجزیه خوشه‌ای برای گروه‌بندی ارقام چای به‌روش حداقل واریانس وارد صحیح بوده است. در تحقیقی که جینز و همکاران (۷) روی داده‌های مزرعه‌ای ذرت انجام دادند، ژنوتیپ‌های مورد مطالعه را با تجزیه خوشه‌ای در ۵ گروه دسته‌بندی کردند و با تابع تشخیص به‌روش کنارگذاری نشان دادند که ۸۰ درصد از گروه‌بندی‌ها صحیح انجام گرفته است. دانایی و همکاران (۲) با بررسی ۱۶ صفت مورفولوژیکی و زراعی در ۴۰۰ ژنوتیپ مختلف سویای ایران، با تجزیه خوشه‌ای به‌روش UPGMA ارقام مورد مطالعه را در ۱۰ گروه مختلف گروه‌بندی کردند و برای تأیید گروه‌بندی به‌دست آمده، از تابع تشخیص استفاده نمودند و نشان دادند که دقت تجزیه خوشه‌ای در انتساب افراد به گروه‌های واقعی ۹۳/۵ درصد بوده است. صبوری و همکاران (۳) نیز ۲۷ رقم برنج را در شرایط مختلف اسمزی مورد مطالعه قرار دادند و آنها را بر اساس تجزیه خوشه‌ای در چهار گروه دسته‌بندی نمودند. نتایج حاصل از تجزیه خوشه‌ای با تجزیه تابع تشخیص مورد بررسی قرار گرفت و مشخص شد که صحت گروه‌بندی ارقام مورد بررسی با تجزیه خوشه‌ای صد درصد بوده است. ابرشهر (۱)۳۰ ژنوتیپ برنج را با روش‌های مختلف تجزیه خوشه‌ای گروه‌بندی نمود و سپس میزان صحت گروه‌بندی‌های حاصل را با تابع تشخیص خطی فیشر مورد ارزیابی قرار داد و به این نتیجه رسید که روش حداقل واریانس "وارد" توانسته است صددرصد ژنوتیپ‌ها را به‌طور صحیح گروه‌بندی نماید. صفایی چایی‌کار (۴) نیز معیارهای مختلف فاصله و روش‌های مختلف تجزیه خوشه‌ای را با استفاده از تجزیه تابع تشخیص خطی فیشر برای گروه‌بندی ۴۹ رقم برنج مورد ارزیابی قرار داد و فاصله اقلیدسی را بهترین معیار برای تعیین فاصله بین ارقام برنج گزارش نمود. هم‌چنین در مقایسه بین روش‌های مختلف تجزیه خوشه‌ای، به‌ترتیب روش‌های UPGMA و حداقل واریانس "وارد" را بهترین روش‌ها معرفی کرد.

در این تحقیق از داده‌های فنوتیپی ۱۴ صفت مهم در ۸ ژنوتیپ کلزا، برای انجام گروه‌بندی اولیه ژنوتیپ‌ها با روش‌های مختلف تجزیه خوشه‌ای، معیارهای متفاوت فاصله و روش‌های مختلف استاندارد کردن داده‌ها استفاده شد. سپس میزان صحت گروه‌بندی‌های حاصل با تجزیه تابع تشخیص مورد بحث و بررسی قرار گرفت.

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

مواد و روش‌ها

تعداد ۸ ژنوتیپ کلزا به نام‌های Hyola60، Hyola308، Hyola330، Hyola401، Hyola420، Option500، RGSS003 و Syn-3 در سال زراعی ۸۵-۱۳۸۴ در قالب طرح بلوک‌های کامل تصادفی با سه تکرار در مؤسسه تحقیقات برنج کشور واقع در شهرستان رشت مورد بررسی قرار گرفتند. عرض هر کرت ۱/۵ متر و طول آن ۳ متر انتخاب شد. کاشت به‌صورت ردیفی و فاصله ردیف‌ها از یک‌دیگر ۲۵ سانتی‌متر بود. کاشت در ۲۶ آبان ماه ۱۳۸۴صورت گرفت و کلیه عملیات زراعی از قبیل مبارزه با علف‌های هرز، مبارزه با آفات و کودپاشی مطابق روش متداول در منطقه انجام شد. یک هفته بعد از سبز شدن بذور و در مرحله چهار برگی، تنک‌کاری و وجین مزرعه صورت گرفت. در طول دوره‌ی رشد و در زمان‌های مناسب، ارزیابی‌های لازم برای ۱۴ صفت شامل تعداد روز از کاشت تا روزت (تعداد روز از زمان کاشت تا موقعی که ۸۰ درصد از بوته‌ها ی هر کرت در روزت باشند)، تعداد روز از روزت تا گل‌دهی (تعداد روز از زمان روزت تا موقعی که ۱۰ درصد از بوته‌ها ی هر کرت به گل رفته باشند)، تعداد روز از گل‌دهی تا رسیدگی (تعداد روز از گل‌دهی تا موقعی که میانگین رطوبت بذور در هر کرت به حدود ۲۰ درصد رسیده باشد)، تعداد برگ ساقه آغوش، تعداد شاخه‌های فرعی درجه اول، تعداد شاخه‌های فرعی درجه دوم، تعداد خورجین در ساقه اصلی، تعداد خورجین در ساقه فرعی درجه اول، تعداد خورجین در ساقه فرعی درجه دوم، ارتفاع بوته، ارتفاع اولین غلاف از سطح خاک، طول خورجین، تعداد دانه در خورجین و عملکرد دانه (کل بوته‌های هر کرت پس از حذف یک ردیف حاشیه از هر طرف کرت برداشت شده و سپس دانه‌های آنها با رطوبت ۱۰ درصد وزن شد) انجام گرفت. ارزیابی صفات دیگر روی ۱۰ بوته در هر کرت که به‌طور تصادفی انتخاب شده بودند، انجام شد. قبل از انجام اندازه‌گیری‌ها، بوته‌های خارج از تیپ حذف و سپس اندازه‌گیری صفات انجام شد و میانگین مشاهدات صفات مختلف در هر کرت جهت انجام تجزیه‌های آماری مورد استفاده قرار گرفت. برای تجزیه داده‌ها، ابتدا تجزیه واریانس طرح بلوکی انجام شد و سپس مقایسه میانگین بین ژنوتیپ‌ها با آزمون توکی صورت گرفت. برای ارزیابی میزان تنوع بین ژنوتیپ‌ها، ضریب تغییرات فنوتیپی و ژنوتیپی هر یک از صفات مورد مطالعه در جمعیت محاسبه شد. برای گروه‌بندی ژنوتیپ‌ها از تجزیه خوشه‌ای بر مبنای میانگین داده‌های هر ژنوتیپ استفاده شد. برای این منظور، ابتدا از داده‌های اصلی و استاندارد نشده استفاده گردید. سپس داده‌ها به سه روش آماری (روابط ۱، ۲ و ۳) استاندارد شدند و به این ترتیب، چهار گروه از داده‌ها برای انجام تجزیه خوشه‌ای به‌کار رفتند.

$$z = \frac{x_i - \bar{x}}{s_x} \qquad [۱]$$

$$\alpha_i^* = \frac{x_i}{x_{Max} - x_{Min}} \qquad [۲]$$

$$\alpha_i^* = \frac{x_i - x_{Min}}{x_{Max} - x_{Min}} \qquad [۳]$$

از معیارهای فاصله اقلیدسی، پیرسون، مینکوسکی، چبی‌چف، کوساین و سیتی بلوک برای تعیین فاصله بین ژنوتیپ‌ها استفاده شد (۱۲). جهت ارزیابی روش‌های مختلف تجزیه خوشه‌ای نیز از هفت روش تجزیه مختلف شامل متوسط فاصله بین کلاسترها (UPGMA) (Un-Weighted Pair Group Using Arithmetic Average)، دورترین همسایه‌ها (Farthest Neighbor)، متوسط فاصله بین و درون کلاسترها (W-Average)، نزدیک‌ترین همسایه‌ها (Nearest Neighbor)، مرکزی (Central)، میانه‌ای (Median) و حداقل واریانس "وارد" (War's Minimum Variance) استفاده شد (۱۲). به‌منظور تعیین تعداد خوشه‌ها نیز از روش بیشترین گسیختگی بر اساس تغییر ناگهانی در اختلاف دو فاصله ادغام متوالی (Fusion value) استفاده گردید (۹ و ۱۲). به این ترتیب که تفاوت مقادیر ادغام گروه‌ها در هر مرحله از تجزیه خوشه‌ای (α_{i+1}) از مقدار قبلی خود (α_i) محاسبه ($\Delta\alpha$) و در هر مرحله از تجزیه که این مقدار تفاوت بیشتری نسبت به سایر مراحل داشت، به‌عنوان نقطه برش دندروگرام انتخاب و تعداد

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

خوشه‌ها بر مبنای آن مشخص شدند (رابطه ۴).

$$\Delta\alpha = \alpha_{i+1} - \alpha_i \qquad [۴]$$

در این رابطه، $\alpha_i = ۱,۲,...,n-۱$، iامین فاصله ادغام در دندروگرام حاصل از تجزیه خوشه‌ای بوده و n تعداد ژنوتیپ‌ها است. در نهایت برای تشخیص صحیح‌ترین گروه‌بندی حاصل از روش‌های مختلف تجزیه خوشه‌ای از روش تجزیه تابع تشخیص به‌روش خطی فیشر (۹ و ۱۳) استفاده شد. برای محاسبه میزان انتساب اشتباه افراد به هر یک از گروه‌ها نیز از روش u- اعتباری استفاده گردید. در این روش در هر بار انجام تجزیه تشخیص، تنها یک عضو از یک گروه کنار گذارده شده و تجزیه تابع تشخیص بر اساس سایر مشاهدات همه‌ی گروه‌ها انجام می‌شود. سپس تفاوت میانگین مشاهدات فرد کنار گذارده شده بر اساس تابع تشخیص حاصل از میانگین گروه‌ها به‌دست می‌آید و فرد به گروهی که به میانگین آن نزدیکتر است، منتسب می‌شود. اگر این فرد به گروه خود منتسب شد، به فراوانی‌های انتساب صحیح یک واحد اضافه می‌شود، در غیر این صورت به فراوانی‌های غیرصحیح یک واحد اضافه خواهد شد. این عمل برای تمامی افراد انجام شد و در پایان از تقسیم فراوانی‌های غیرصحیح بر تعداد کل افراد احتمال انتساب اشتباه و از تقسیم فراوانی‌های صحیح بر تعداد کل افراد احتمال انتساب صحیح براورد شد. پس از انجام تجزیه تشخیص و برای تأیید نتایج حاصل و تعیین تعداد واقعی گروه‌ها، از T^2 هتلینگ، CCC پلات و تجزیه واریانس چند متغیره (۹) استفاده گردید. تجزیه واریانس و مقایسه میانگین ژنوتیپ‌ها با نرم‌افزار SAS نسخه ۶/۱۲ انجام شد. سایر تجزیه‌های آماری شامل تجزیه واریانس چند متغیره، تجزیه خوشه‌ای، تجزیه تابع تشخیص، T^2 هتلینگ و CCC پلات با نرم افزار SPSS نسخه ۱۴ انجام شدند.

نتایج و بحث

تجزیه واریانس داده‌ها اختلاف معنی‌داری را بین ژنوتیپ‌ها از نظر کلیه صفات مورد مطالعه نشان داد. مقایسه میانگین بین ژنوتیپ‌ها نشان داد که ژنوتیپ Hyola401 از نظر عملکرد دانه و بسیاری از صفات دیگر برتر از سایر ژنوتیپ‌ها بود (جدول ۱). از آنجایی که در استان گیلان، کلزا به‌عنوان کشت دوم مطرح می‌باشد، بنابراین ژنوتیپی که بتواند علاوه بر تولید عملکرد دانه بالا، طول دوره رشد کوتاه‌تری داشته باشد تا با کشت اصلی یعنی برنج تداخل نکند، از اهمیت زیادی برخوردار است. نتایج مقایسه میانگین بین ژنوتیپ‌ها از نظر صفات مرتبط با طول دوره رشد نشان داد که ژنوتیپ Hyola308 از نظر سه صفت تعداد روز از کاشت تا روزت، تعداد روز از روزت تا گل‌دهی و تعداد روز از گل‌دهی تا رسیدگی دارای کمترین طول دوره رشد در بین ژنوتیپ‌های مورد مطالعه بود، اما کمترین مقدار عملکرد دانه را در بین کلیه ژنوتیپ‌ها داشت. در مقابل ژنوتیپ Hyola401 نه تنها بیشترین عملکرد دانه را در بین کلیه ژتوتیپ‌ها به خود اختصاص داد، بلکه از نظر دوره رشد نیز در حد بسیار خوبی قرار داشت، به‌طوری که از نظر هر سه صفت مربوط به‌طول دوره رشد تفاوت معنی‌داری با Hyola308 نداشت. بنابراین این ژنوتیپ می‌تواند برای استفاده به‌عنوان کشت دوم در شمال کشور مناسب باشد. ضریب تغییرات فنوتیپی (CV_p) و ژنوتیپی (CV_g) صفات مختلف در جدول ۲ ارائه شده است. میزان تنوع بین ژنوتیپ‌ها از نظر بیشتر صفات بالا بود (جدول ۲) و نشان داد که با انجام گزینش می‌توان ژنوتیپ‌های مناسب را انتخاب و ارزش فنوتیپی و ژنوتیپی جمعیت را افزایش داد. صفت تعداد روز از کاشت تا روزت و تعداد شاخه‌های فرعی درجه دوم به‌ترتیب با ۴/۳۲ و ۵۱/۶۷ درصد کمترین و بیشترین ضریب تغییرات فنوتیپی را داشتند. ضرایب تغییرات ژنوتیپی این دو صفت نیز به‌ترتیب ۴/۰۱ درصد و ۴۸/۷۴ درصد بود و به‌ترتیب کمترین و بیشترین تنوع ژنوتیپی را در بین صفات دارا بودند. پس از این دو صفت، عملکرد دانه به‌ترتیب با ۳۱/۵۴ و ۳۱/۲۸ درصد دارای بیشترین ضریب تغییرات فنوتیپی و ژنوتیپی بود. از آنجایی که برای اکثر صفات، بین میزان ضریب تغییرات ژنوتیپی و فنوتیپی اختلاف چندانی مشاهده نشد، بنابراین انتخاب

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

جدول۱. مقایسه میانگین ژنوتیپ‌های کلزا از نظر صفات مورد مطالعه

تیمارها	صفات مورد مطالعه													
	عملکرد دانه	تعداد دانه در خوردجین	طول خوردجین	تعداد خوردجین در ساقه فرعی درجه دوم	تعداد خوردجین در ساقه فرعی درجه اول	تعداد خوردجین در ساقه اصلی	تعداد ساقه فرعی درجه دوم	تعداد ساقه فرعی درجه اول	ارتفاع اولین خوردجین از خاک	ارتفاع بوته	تعداد برگ ساقه آغوش	تعداد روز از گل‌دهی تا رسیدگی	تعداد روز از روزت تا گل‌دهی	تعداد روز از کاشت تا روزت
Hyola401	۳/۶۱^{a}	۲۸^{a}	۵/۶۶^{ab}	۱۱/۶۷^{b}	۳۷^{a}	۴۶/۶۷^{cd}	۱۰/۸۳^{a}	۷/۰۷^{a}	۶۳/۸۳^{ab}	۱۳۷/۳^{bc}	۴/۵^{b}	۷۲^{b}	۳۵^{b}	۷۱^{b}
Hyola420	۳/۲۸^{b}	۲۰/۶۷^{c}	۵/۸۳^{ab}	۱۵^{ab}	۲۲/۶۷^{cd}	۳۸/۶۷^{d}	۵^{bc}	۴/۲۷^{d}	۶۱/۶۷^{ab}	۱۲۴/۱۵^{de}	۳/۵۳^{d}	۷۷/۶۷^{a}	۳۹/۳۳^{ab}	۷۲^{b}
Syn-3	۲/۵۱^{c}	۲۴/۹۷^{b}	۵/۸۴^{ab}	۱۴^{ab}	۲۰/۳۳^{d}	۴۲^{d}	۶/۳۳^{b}	۶/۴۷^{ab}	۷۱/۱۷^{ab}	۱۳۷/۶۶^{b}	۷/۵۳^{a}	۷۰^{b}	۳۶/۶۷^{ab}	۷۲^{b}
Hyola60	۱/۸۶^{e}	۱۶/۳^{d}	۶/۱۶^{a}	۱۴^{ab}	۲۹^{b}	۵۴/۶۷^{bc}	۴/۶۷^{bc}	۵/۴^{bcd}	۷۶/۵^{a}	۱۲۹/۲۴^{cd}	۴/۵۷^{c}	۶۹^{b}	۴۲^{a}	۷۲^{b}
Hyola308	۱/۶۳^{e}	۲۰/۶^{c}	۶/۲۶^{a}	۱۶/۶۷^{a}	۲۶^{bc}	۴۵^{d}	۲^{c}	۴/۲۷^{d}	۵۵/۸۳^{b}	۱۰۲/۲۲^{f}	۴/۵^{c}	۶۷^{b}	۳۳/۶۷^{b}	۷۰/۳۳^{b}
Option500	۲/۱۸^{d}	۲۶/۸۷^{ab}	۶/۱۸^{a}	۱۷^{a}	۳۵/۳۳^{a}	۶۱^{ab}	۵/۸۳^{bc}	۶/۴^{abc}	۶۴/۶۷^{ab}	۱۵۲/۶۳^{a}	۷/۹^{a}	۷۷/۳۳^{a}	۴۱/۳۳^{ab}	۷۹/۶۷^{a}
Hyola330	۳/۱۱^{b}	۲۰/۷۶^{c}	۵/۱۳^{b}	۱۲^{b}	۳۸/۳۳^{a}	۶۵/۶۷^{a}	۴/۸۳^{bc}	۵/۱۷^{cd}	۶۶^{ab}	۱۱۷/۹۷^{e}	۴/۹^{bc}	۷۰/۳۳^{b}	۳۸/۶۷^{ab}	۷۲/۷۶۷^{b}
RGSS003	۱/۶۳^{e}	۲۸/۹۷^{a}	۵/۹۶^{a}	۱۱/۶۷^{b}	۳۸/۶۷^{a}	۶۷/۶۷^{a}	۱۱/۱۷^{a}	۶/۸۳^{a}	۶۷/۸۳^{ab}	۱۳۲/۵۳^{bc}	۵/۶^{b}	۷۰^{b}	۴۱/۶۷^{a}	۷۰/۶۷^{b}

جدول۲. ضریب تغییرات فنوتیپی و ژنوتیپی جمعیت مورد مطالعه از نظر صفات مختلف

صفات	تعداد روز از کاشت تا روزت	تعداد روز از روزت تا گل‌دهی	تعداد روز از گل‌دهی تا رسیدگی	تعداد برگ ساقه آغوش	تعداد شاخه‌های فرعی درجه اول	تعداد شاخه‌های فرعی درجه دوم	تعداد خورجین در ساقه اصلی
ضریب تغییرات فنوتیپی (درصد)	۴/۳۲	۸/۷۵	۶/۵۲	۲۷/۶۷	۲۰/۱۱	۵۱/۶۷	۲۱/۵۱
ضریب تغییرات ژنوتیپی (درصد)	۴/۰۱	۶/۸۰	۶/۳۸	۲۷/۳۷	۱۹/۱۱	۴۸/۷۴	۲۱/۰۳

ادامه جدول ۲

صفات	تعداد خورجین در ساقه فرعی درجه اول	تعداد خورجین در ساقه فرعی درجه دوم	ارتفاع بوته	ارتفاع اولین غلاف از سطح خاک	طول خورجین	تعداد دانه در خورجین	عملکرد دانه
ضریب تغییرات فنوتیپی (درصد)	۲۳/۷۱	۱۶/۹۴	۱۱/۵۶	۱۱/۱۸	۶/۸۱	۱۹/۳۸	۳۱/۵۴
ضریب تغییرات ژنوتیپی (درصد)	۲۳/۳۲	۱۴/۳۶	۱۱/۴۷	۸/۳۷	۵/۸۹	۱۹/۱۹	۳۱/۲۸

ژنوتیپ‌ها بر اساس ارزش‌های فنوتیپی آنها می‌تواند معیار مناسبی برای گزینش باشد.

تجزیه خوشه‌ای با استفاده از داده‌های حاصل از روش‌های مختلف استاندارد کردن داده‌ها و داده‌های استاندارد نشده با روش‌های حداقل واریانس "وارد" و UPGMA نشان داد که معیارهای فاصله اقلیدسی و مینکوسکی ژنوتیپ‌ها را در سه گروه، معیارهای فاصله پیرسون، کوساین و سیتی بلوک ژنوتیپ‌ها را در چهار گروه و معیار فاصله چبی‌چف ژنوتیپ‌ها را در پنج گروه قرار دادند (جدول ۳). مقایسه دندروگرام‌های حاصل با استفاده از تجزیه تابع تشخیص خطی فیشر حاکی از آن بود که معیار فاصله اقلیدسی با استفاده از داده‌های استاندارد شده با هر سه روش توانست ژنوتیپ‌ها را با احتمال صحت گروه‌بندی ۸۷/۵ درصد بهتر از سایر معیارهای فاصله گروه‌بندی کرده و تفاوت بین ژنوتیپ‌ها را نشان دهد. احتمال انتساب اشتباه ژنوتیپ‌ها به گروهی که متعلق به آن نبودند، در این روش تنها ۱۲/۵ درصد بود. معیار فاصله چبی‌چف نیز در مورد تمامی انواع داده‌ها کمترین احتمال ممکن را در صحیح نسبت دادن ژنوتیپ‌ها به گروه‌های مختلف دارا بود و با احتمال صحت ۲۵ درصد ضعیف‌ترین گروه‌بندی را در تمامی روش‌های تجزیه خوشه‌ای ارائه داد (جدول ۳). این نتیجه علاوه بر تأیید معیار فاصله اقلیدسی به‌عنوان یک معیار فاصله مناسب برای داده‌های مورفولوژیک و کمی، نشان داد که هر سه روش استاندارد کردن داده‌ها نتایج مشابهی ارائه داده و بهتر از استاندارد نکردن داده‌ها بودند.

با توجه به نتایج بهتر معیار فاصله اقلیدسی نسبت به سایر معیارها و هم‌چنین یکسان بودن نتایج روش‌های استاندارد کردن داده‌ها در تجزیه خوشه‌ای در این دو روش، گروه‌بندی ژنوتیپ‌ها به‌روش‌های مختلف تجزیه خوشه‌ای با استفاده از معیار فاصله اقلیدسی و استاندارد کردن داده‌ها به‌روش دوم (رابطه ۲) انجام شد و صحت گروه‌بندی‌ها در روش‌های مختلف به‌کمک تجزیه تابع تشخیص مورد تجزیه و تحلیل قرار گرفت و روش‌های مختلف تجزیه خوشه‌ای با هم مقایسه شدند.

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

جدول۳. جدول نسبت موفقیت افراد درون گروه‌ها با تابع تشخیص

معیار فاصله	گروه	تعداد ژنوتیپ	گروه‌ها و میزان صحت گروه‌بندی ژنوتیپ‌ها در هر گروه با تجزیه تشخیص			
			۱	۲	۳	
دسته اول (معیار فاصله اقلیدسی و مینکوسکی)	۱	۳	۳ ٪۱۰۰			
	۲	۴		۴ ٪۱۰۰		
	۳	۱		۱ ٪۰	۰	
		صحت گروه‌بندی			٪۸۷/۵	
	گروه	تعداد ژنوتیپ	۱	۲	۳	۴
دسته دوم (معیار فاصله پیرسون، کوساین و سیتی‌بلوک)	۱	۲	۲ ٪۱۰۰			
	۲	۴		۴ ٪۱۰۰		
	۳	۱		۱	۰ ٪۰	
	۴	۱		۱		۰ ٪۰
		صحت گروه‌بندی			٪۷۵	
	گروه	تعداد ژنوتیپ	۱	۲	۳	۴	۵
دسته سوم (معیار فاصله چبی‌چف)	۱	۲	۰ ٪۰	۱	۱		
	۲	۲	۰	۰ ٪۰	۱	۱	
	۳	۲			۲ ٪۱۰۰	۰	
	۴	۱			۱	۰ ٪۰	
	۵	۱		۱			۰ ٪۰
		صحت گروه‌بندی				٪۲۵	

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

دندروگرام حاصل از تجزیه خوشه‌ای برای روش‌های مختلف در شکل‌های ۱ تا ۷ نشان داده شده است. نقطه برش بر اساس تغییر ناگهانی در اختلاف دو فاصله ادغام متوالی بدست آمد و دندروگرام‌های حاصل در آن نقاط برش داده شدند. بر اساس برش‌های ایجاد شده در دندروگرام‌ها، این روش‌ها در سه دسته قرار گرفتند که روش‌های هر دسته گروه‌بندی مشابهی داشتند. دسته اول شامل روش‌های UPGMA، دورترین همسایه‌ها و حداقل واریانس "وارد" بود. این سه روش، ژنوتیپ‌ها را در سه گروه قرار دادند و نتایج مشابهی داشتند. دسته دوم شامل روش‌های متوسط فاصله بین و درون کلاسترها (W-Average)، مرکزی و میانه‌ای بود و ژنوتیپ‌ها را در ۴ گروه دسته‌بندی کردند. دسته سوم شامل تنها روش نزدیک‌ترین همسایه‌ها بود و ژنوتیپ‌ها را در ۳ گروه قرار داد، اما گروه‌بندی حاصل با روش‌های دسته اول متفاوت بود (شکل‌های ۱ تا ۷).

برای ارزیابی گروه‌بندی حاصل از این روش‌ها، ضرایب هم‌بستگی کوفنتیک بین ماتریس‌های ورودی فاصله و خروجی دندروگرام محاسبه گردید که برای روش‌های UPGMA، دورترین همسایه‌ها، متوسط فاصله بین و درون کلاسترها، نزدیک‌ترین همسایه‌ها ، مرکزی، میانه‌ای و حداقل واریانس "وارد" به‌ترتیب برابر با ۰/۶۸، ۰/۶۴، ۰/۶۵، ۰/۵۷، ۰/۶۱، ۰/۶۶ و ۰/۵۹ بود. با توجه به پایین بودن ضریب هم‌بستگی کوفنتیک در تمامی این روش‌ها نمی‌توان گفت که کدام روش بهتر بوده و تجزیه خوشه‌ای را به‌نحو مطلوبی انجام داده است. ولی با توجه به دندروگرام‌ها می‌توان مشاهده کرد که در اکثر روش‌ها (به غیر از روش‌های UPGMA، دورترین همسایه‌ها و حداقل واریانس وارد) حالت زنجیره‌ای یا پله‌ای مشاهده می‌شود، به‌عبارت دیگر در اولین مرحله از تجزیه خوشه‌ای دو ژنوتیپ در هم ادغام شده و در یک گروه قرار گرفته‌اند و سپس در مراحل بعدی یک ژنوتیپ به گروه قبلی اضافه شده و حالت زنجیره‌ای ایجاد نموده‌اند که یک وضعیت نامطلوب در گروه‌بندی ژنوتیپ‌ها است (شکل‌های ۱ الی ۷). به این ترتیب، روش‌های مختلف تجزیه خوشه‌ای که در ۳ دسته با گروه‌بندی‌های متفاوت قرار گرفتند، به‌وسیله تجزیه تابع تشخیص مورد ارزیابی قرار گرفته و صحت گروه‌بندی آنها مورد بررسی قرار گرفت. علاوه بر آن، تعداد واقعی گروه‌ها با تجزیه واریانس چند متغیره، T^2 هتلینگ و پلات CCC مورد بررسی قرار گرفت.

توابع تشخیص به‌دست آمده برای روش‌های دسته اول که ژنوتیپ‌ها را در سه گروه قرار دادند و شامل روش‌های UPGMA، حداقل واریانس وارد و دورترین همسایه‌ها بودند، به این شرح بود:

$$B_1 = ۲/۴۹۶X_2 + ۲/۴۳۳X_{13}$$
$$B_2 = -۰/۳۵۷X_2 + ۰/۶۶۲X_{13}$$

که در آن X_2 تعداد روز از روزت تا گل‌دهی و X_{13} تعداد دانه در خورجین می‌باشد و سایر متغیرها هم به‌دلیل معنی‌دار نبودن ضرایب آنها در معادله وارد نشدند.

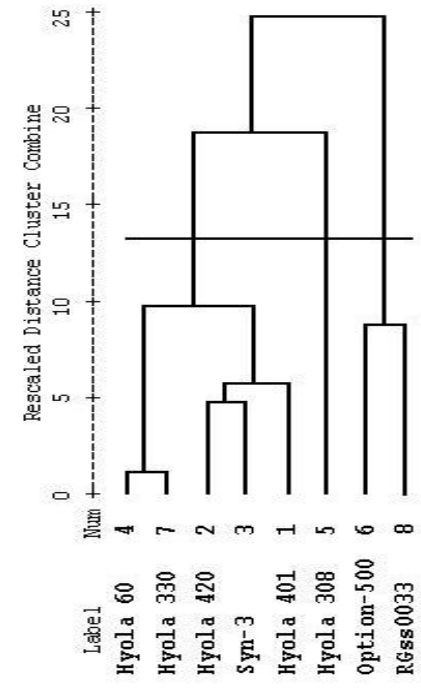
تجزیه تابع تشخیص خطی فیشر برای آزمون صحت گروه‌بندی اولیه ژنوتیپ‌ها در این دسته با استفاده از روش u– اعتباری نشان داد که از بین سه گروه ایجاد شده، دو گروه با احتمال ۱۰۰ درصد صحیح و یک گروه به‌طور کاملاً نادرست گروه‌بندی شده است. در نهایت هر سه روش موجود در این دسته با احتمال ۸۷/۵ درصد توانستند ژنوتیپ‌ها را صحیح گروه‌بندی نمایند. با مقایسه بین گروه‌بندی اولیه و گروه‌بندی حاصل از تابع تشخیص می‌توان مشاهده کرد که تابع تشخیص این ژنوتیپ‌ها را در دو گروه قرار داد و ژنوتیپ Hyola308 به‌جای قرار گرفتن در یک گروه مجزا، باید در گروه اول و در کنار ژنوتیپ‌های Hyola60، Hyola330، Hyola401، Hyola420 و Syn-3 قرار گیرد (شکل۸).

توابع تشخیص به‌دست آمده برای روش‌های دسته دوم که شامل روش‌های متوسط فاصله بین و درون کلاسترها، مرکزی و میانه‌ای بود و ژنوتیپ‌ها را در چهار گروه قرار دادند، به شرح زیر بود:

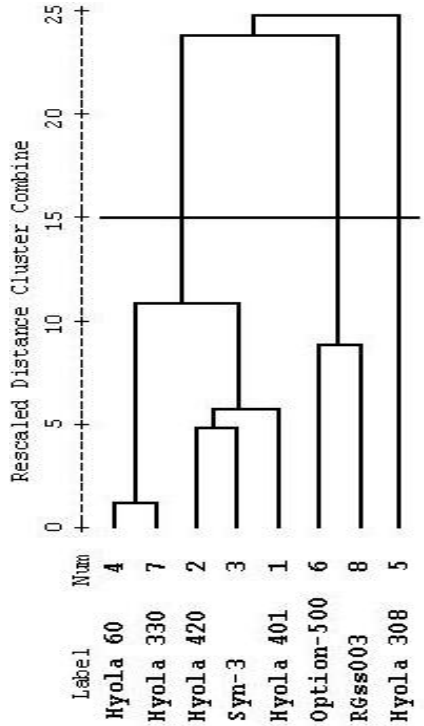
$$B_1 = ۲/۱۶۲X_1 + ۱/۹۲۷X_7$$
$$B_2 = -۰/۰۲۱X_1 + ۰/۹۸۱X_7$$

که در آن X_1 تعداد روز از کاشت تا روزت و X_7 تعداد

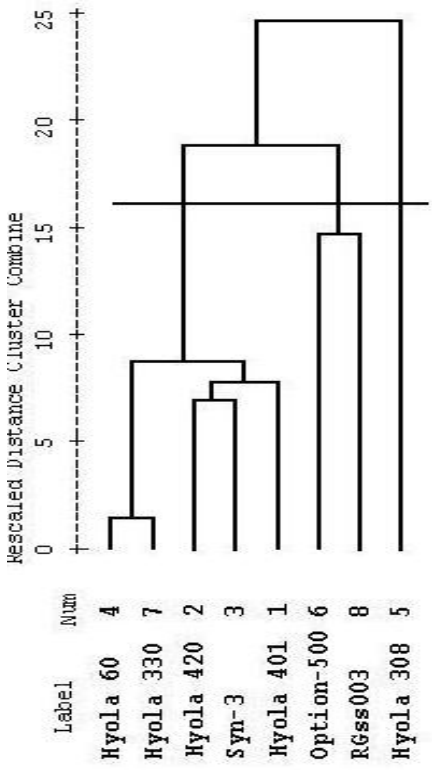
[DOR: 20.1001.1.24763594.1388.13.47.43.1]



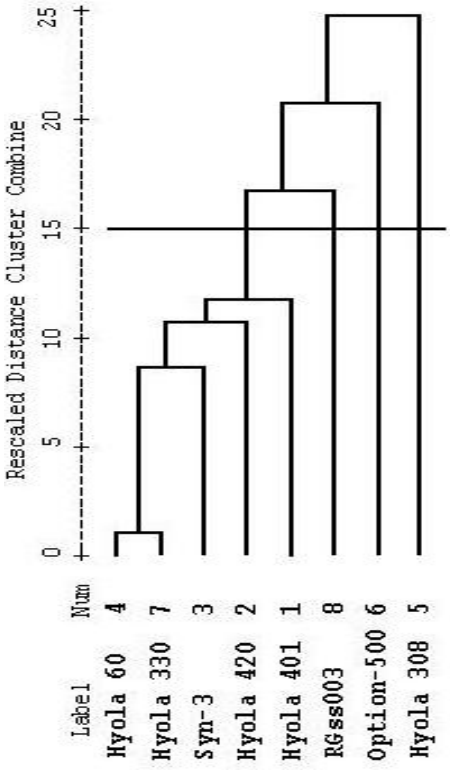
شکل۱. دندروگرام روش حداقل واریانس وارد



شکل۲. دندروگرام روش دورترین همسایه‌ها

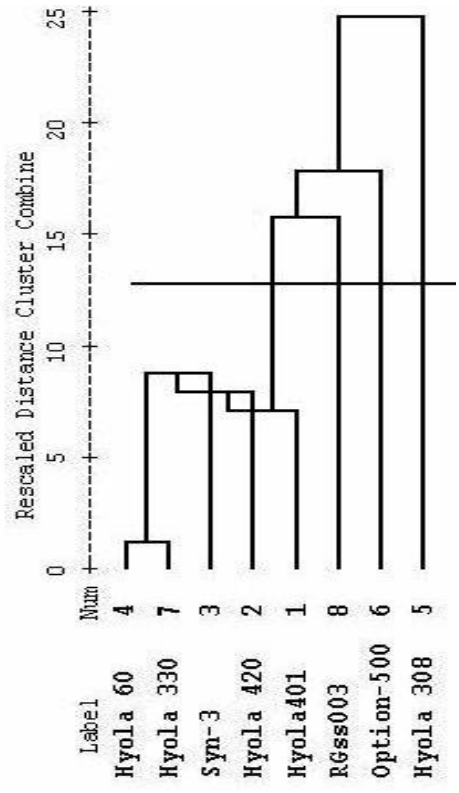


شکل۳. دندروگرام روش متوسط فاصله بین کلاسترها (UPGMA)

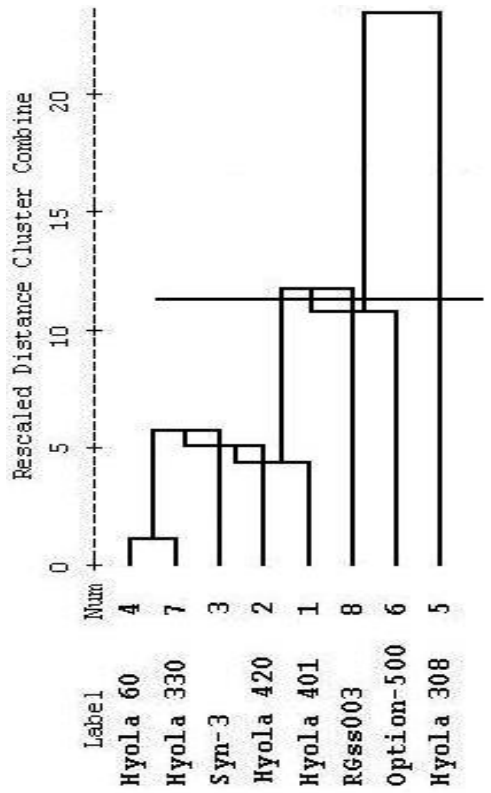


شکل۴. دندروگرام روش متوسط فاصله بین و درون کلاسترها (W-Average)

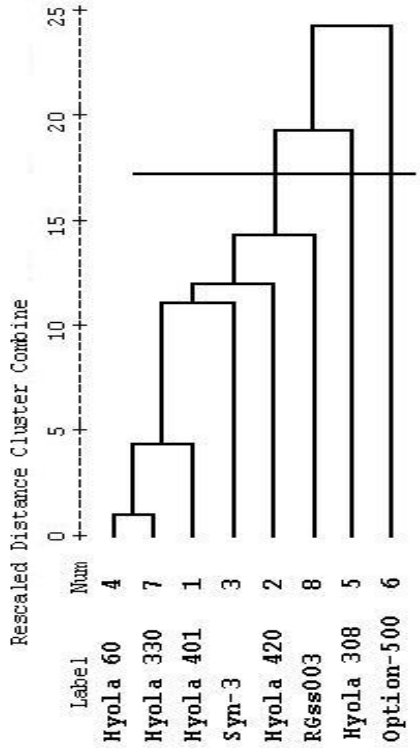
[DOR: 20.1001.1.24763594.1388.13.47.43.1]



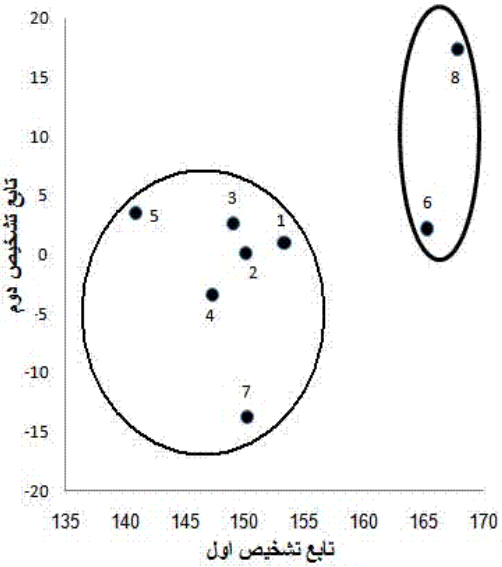
شکل۵. دندروگرام روش مرکزی



شکل۶. دندروگرام روش میانه‌ای



شکل۷. دندروگرام روش نزدیک‌ترین همسایه‌ها



شکل۸. گروه‌بندی ژنوتیپ‌ها بر اساس تابع تشخیص برای روش‌های دسته اول

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

شاخه فرعی درجه دوم می‌باشد و سایر متغیرها هم به‌دلیل معنی‌دار نبودن ضرایب آنها در معادله وارد نشدند. نتایج نشان داد که از بین ۴ گروه ایجاد شده به‌وسیله روش‌های این دسته، تنها یک گروه با احتمال ۸۰ درصد صحیح گروه‌بندی شده و سه گروه دیگر کاملاً نادرست گروه‌بندی شده بودند، به‌طوری که احتمال انتساب اشتباه برای این گروه‌بندی‌ه ۵۰ درصد به‌دست آمد. هم‌چنین نتایج نشان داد که ژنوتیپ‌ها به‌جای قرار گرفتن در ۴ گروه باید در ۲ گروه قرار گیرند (شکل ۹).

گروه‌های به‌دست آمده با روش نزدیک‌ترین همسایه‌ها نیز با روش u- اعتباری مورد ارزیابی قرار گرفتند. نتایج نشان داد که از بین سه گروه ایجاد شده، تنها یک گروه ۱۰۰ درصد صحیح گروه‌بندی شده و دو گروه دیگر کاملاً نادرست گروه‌بندی شده‌اند. محاسبه احتمال انتساب اشتباه ژنوتیپ‌ها نیز نشان داد که این روش توانسته است با احتمال ۷۵ درصد ژنوتیپ‌ها را صحیح گروه‌بندی نماید. توابع تشخیص براورد شده برای این روش هم به‌صورت زیر بود:

$$B_1 = ۱۲/۳۹۷X_1 + ۷/۱۶۳X_2 + ۱۲/۹۸۱X_5 - ۴/۹۵۳X_{14}$$
$$B_2 = -۰/۲۳۲X_1 + ۱/۱۰۱X_2 + ۰/۰۵۹X_5 + ۱/۰۲۷X_{14}$$

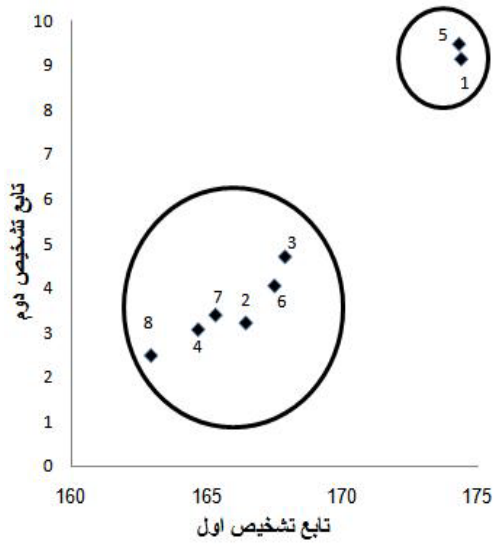
که در آن X_1 تعداد روز از کاشت تا روزت، X_2 تعداد روز از روزت تا گل‌دهی، X_5 ارتفاع بوته و X_{14} عملکرد دانه در هکتار می‌باشد و سایر متغیرها هم به‌دلیل معنی‌دار نبودن ضرایب آنها در معادله وارد نشدند. هم‌چنین نتایج نشان داد که ژنوتیپ‌ها به‌جای قرار گرفتن در ۳ گروه باید در ۱ گروه قرار گیرند (شکل ۱۰).

برای تأیید نتایج حاصل و تعیین تعداد واقعی گروه‌ها، از آزمون‌های T^2 هتلینگ و معیار توان سوم گروه‌ها (پلات CCC) استفاده شد (جدول ۴). بر پایه‌ی آزمون T^2 هتلینگ، با کاهش تعداد گروه‌ها از ۷ به تعداد گروه‌های کمتر، مقدار T^2 که معیاری برای ترکیب دو گروه می‌باشد، به تدریج افزایش پیدا کرد و در ناحیه دو گروه حداکثر مقدار T^2 مشاهده گردید و مجدداً مقدار آن کاهش یافت که این امر نشان داد که تعداد دو گروه، بهترین گروه‌بندی برای ژنوتیپ‌های مورد مطالعه می‌باشد (جدول ۴). تعداد گروه‌ها بر اساس پلات CCC در ناحیه‌ای که این معیار حداکثر و مثبت است، مشخص می‌شود. با توجه به جدول ۴ مشاهده می‌شود که با کاهش تعداد گروه‌ها از ۷ به ۲ گروه، مقدار CCC از مقادیر منفی به بیشترین مقدار مثبت خود (۱/۷۵) رسید و پس از آن یعنی ادغام کلیه‌ی ژنوتیپ‌ها در یک گروه موجب شد که این آماره مجدداً کاهش یافته و به مقدار صفر برسد. به این ترتیب، بر این اساس نیز تعداد دو گروه تعیین شد (جدول ۴).

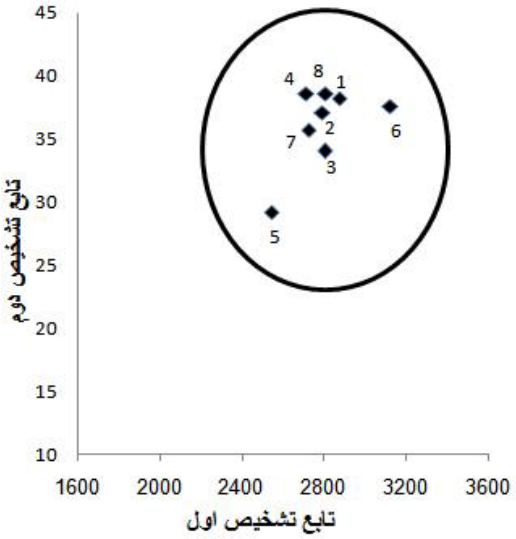
علاوه بر آزمون‌های فوق، از تجزیه واریانس چند متغیره نیز برای آزمون صحت گروه‌بندی حاصل از تجزیه تابع تشخیص استفاده شد (جدول ۵). به این ترتیب که دو گروه پیشنهادی تابع تشخیص (گروه اول شامل ۶ ژنوتیپ Hyola60، Hyola308، Hyola330، Hyola401، Hyola420 و Syn-3 و گروه دوم شامل دو ژنوتیپ Option500 و RGSS003) به‌عنوان دو تیمار و ژنوتیپ‌های درون آنها به‌عنوان تکرار در نظر گرفته شدند و تجزیه واریانس چند متغیره در قالب طرح کاملاً تصادفی نامتعادل انجام شد. نتایج حاصل از هر سه آزمون استفاده شده (جدول ۵) نشان داد که اختلاف بسیار معنی‌داری بین دو گروه از ژنوتیپ‌ها وجود داشت. به این ترتیب، گروه‌بندی حاصل از تابع تشخیص خطی فیشر باز هم مورد تأیید قرار گرفت.

در مقایسه روش‌های مختلف تجزیه خوشه‌ای، نوع داده‌های مورد استفاده و روش‌های مختلف تجزیه خوشه‌ای برای گروه‌بندی ژنوتیپ‌ها می‌توان گفت که فاصله اقلیدسی از سایر معیارهای فاصله بهتر بود، چرا که از مزایای این فاصله این است که اگر ماتریس داده‌ها دارای مقادیر از دست رفته باشد، باز هم می‌توان از این معیار استفاده نمود (۹). هم‌چنین از این معیار به‌نحو مطلوب می‌توان برای ارزیابی فاصله بین ژنوتیپ‌ها به‌ویژه برای صفات مورفولوژیک و کمی استفاده نمود. این موضوع با انجام تجزیه تابع تشخیص در این مقاله نیز به اثبات رسید، به‌طوری که در تمامی روش‌های مورد استفاده، احتمال گروه‌بندی صحیح ژنوتیپ‌ها با این معیار فاصله ۸۷/۵ درصد و

[DOR: 20.1001.1.24763594.1388.13.47.43.1]



شکل۹. گروه‌بندی ژنوتیپ‌ها بر اساس تابع تشخیص برای روش‌های دسته دوم



شکل۱۰. گروه‌بندی ژنوتیپ‌ها بر اساس تابع تشخیص برای روش‌های دسته سوم

جدول۴. آزمون T^2 هتلینگ و معیار توان سوم گروه‌ها (CCC) برای تعیین تعداد واقعی گروه‌ها

تعداد گروه‌ها	T^2 هتلینگ	پلات CCC
۷	۰	-۰/۷
۶	۰	-۱/۵
۵	۱/۳	-۰/۹
۴	۱/۴	۰
۳	۱/۸	۰/۵
۲	۳/۵	۱/۷۵
۱	۲/۸	۰/۰۰۰

جدول۵. خلاصه تجزیه واریانس چند متغیره بین دو گروه حاصل از تجزیه تابع تشخیص بر مبنای طرح کاملاً تصادفی نامتعادل

منابع تغییر	درجه آزادی	آماره		F معادل	سطح معنی‌دار
تیمار	۱	اثر پیلای (Pillai trace)	۶/۸۸۴	۲۹/۷۶	۰/۰۰۰
		ویلکس لامبدا (Wilk's lambda)	۰/۰۰۰	۹۸/۸۶	۰/۰۰۰
		بزرگترین ریشه روی (Roy's greatest root)	۷۳/۵۰۰	۲۹۵/۴۰	۰/۰۰۰
خطای آزمایش	۶	–	–	–	–

بالاتر از سایر معیارهای فاصله بود.

علاوه بر معیار فاصله، نوع داده‌های مورد استفاده در تجزیه خوشه‌ای و گروه‌بندی ژنوتیپ‌ها، یعنی به‌کار بردن داده‌های استاندارد شده و یا استفاده از داده‌های خام اولیه، می‌تواند مطلوبیت گروه‌بندی حاصل را تحت تأثیر قرار دهد. بدیهی است که برای انجام تجزیه خوشه‌ای، صفات مختلفی مورد اندازه‌گیری قرار می‌گیرندکه دارای مقیاس اندازه‌گیری متفاوتی هستند. علاوه بر آن، تنوع یا واریانس این صفات نیز بسیار متنوع است، به‌طوری که اگر مستقیماً از داده‌های خام اولیه برای انجام تجزیه خوشه‌ای استفاده گردد، معیار فاصله مورد استفاده برای ارزیابی میزان فاصله ژنتیکی بین ژنوتیپ‌ها و به‌دنبال آن گروه‌بندی انجام شده کاملاً تحت تأثیر صفات با مقیاس کوچکتر (که دارای اعداد بزرگتری هستند) و یا صفات با واریانس بزرگ‌تر قرار گرفته و لذا عموماً نتایج مطلوبی به‌دست نخواهد آمد (۹ و ۱۲). با استاندارد کردن داده‌ها، اثر صفات با مقیاس کوچکتر و واریانس بزرگ‌تر تعدیل شده و همه صفات دارای وزن یکسانی خواهند شد و لذا در تجزیه خوشه‌ای و تشکیل گروه‌های حاصل سهم یکسانی نیز خواهند داشت. نتایج حاصل از تجزیه تابع تشخیص و براورد احتمال انتساب صحیح ژنوتیپ‌ها به گروه‌ها نیز این موضوع را مورد تأیید قرار داد، به‌طوری که گروه‌بندی حاصل با داده‌های استاندارد شده در تمامی روش‌های تجزیه خوشه‌ای کاملاً متفاوت از گروه‌های حاصل از داده‌های استاندارد نشده بود. اما تفاوتی بین نتایج حاصل از داده‌های استاندارد شده در سه روش (روابط ۱، ۲ و۳) وجود نداشت وتجزیه تابع تشخیص، گروه‌بندی حاصل از هر ۳ مجموعه داده‌های استاندارد شده را کاملاً مشابه و صحت آنها را ۸۷/۵ درصد محاسبه نمود.

روش تجزیه خوشه‌ای مورد استفاده نیز مستقیماً می‌تواند گروه‌بندی‌های حاصل را تحت تأثیر قرار دهد. به عقیده بسیاری از محققین، روش‌های متوسط فاصله بین کلاسترها و حداقل واریانس وارد نتیجه بهتری از سایر روش‌ها ارئه می‌دهند (۹). روش متوسط فاصله بین کلاسترها از متوسط فاصله بین ژنوتیپ‌ها یا گروه‌ها استفاده کرده و دو گروهی که متوسط فاصله بین آنها کمترین مقدار را داشته باشد در هم ادغام می‌شوند. در مقابل، روش حداقل واریانس وارد بر مبنای محاسبه مجموع مربعات درون گروهی استوار بوده و لذا فاصله بین گروه‌ها یا ژنوتیپ‌ها را بزرگتر کرده و بهتر نشان می‌دهد. به این ترتیب یک نوع معیار بهینگی همانند روش‌های تجزیه براورد می‌شود، اگرچه این معیار بهینگی مطلق نیست. بنابراین به‌نظر می‌رسد این روش خصوصاً برای صفات کمی و مزرعه‌ای گروه‌بندی بهتری ایجاد نماید. مطلوب بودن روش‌های متوسط فاصله بین کلاسترها و حداقل واریانس وارد با تجزیه تابع تشخیص نیز به اثبات رسید به‌طوری که احتمال انتساب صحیح ژنوتیپ‌ها به گروه‌های مربوطه ۸۷/۵ درصد براورد شد. علاوه بر آن، دندروگرام حاصل از هر دو روش فوق، گروه‌بندی بسیار مطلوبی داشته و با خصوصیات مورفولوژیک و کمی ژنوتیپ‌های مورد مطالعه کاملاً مطابقت داشت.

تجزیه تابع تشخیص برای آزمون صحت گروه‌بندی حاصل از تجزیه خوشه‌ای توسط محققین دیگر نیز بررسی شده است (۱، ۲، ۳، ۴، ۵، ۷ و ۱۱). موردا و همکاران (۱۱) با انجام تجزیه خوشه‌ای به روش حداقل واریانس وارد و معیار فاصله اقلیدسی، ۸۵ رقم چای را در دو گروه آسیایی و آفریقایی گروه‌بندی نمودند و سپس با تجزیه تابع تشخیص نشان دادند که ۹۴/۴ درصد از این گروه‌بندی، صحیح انجام شده است. هم‌چنین جینز و همکاران (۷) با تابع تشخیص به‌روش کنارگذاری نشان دادند که ۸۰٪ از گروه‌بندی ارقام ذرت با تجزیه خوشه‌ای صحیح بوده است. بالوچی و همکاران (۵) نیز مطالعه‌ای را بر روی صفات زراعی ۱۲۵ اکوتیپ گیاه بروموس مناطق مختلف انجام دادند و با تجزیه خوشه‌ای آنها را در ۴ گروه دسته‌بندی کردند. سپس با تابع تشخیص نشان دادند که ۹۴/۴ درصد از این اکوتیپ‌ها درست و ۵/۶ درصد از آنها (۷ اکوتیپ) نادرست گروه‌بندی شده بودند. ابرشهر (۱)، صبوری و همکاران (۳) و صفایی چایی‌کار (۴) نیز نتایج حاصل از تجزیه خوشه‌ای در گروه‌بندی ارقام برنج را با تجزیه تابع

[DOR: 20.1001.1.24763594.1388.13.47.43.1]

تشخیص ارزیابی و صحت گروه‌بندی‌های حاصل را صد در صد عنوان نمودند.

نتیجه‌گیری

نتایج حاصل از این آزمایش نشان داد که انتخاب روش صحیح تجزیه خوشه‌ای و معیار فاصله مورد استفاده در بررسی تنوع ژنتیکی و گروه‌بندی ژنوتیپ‌ها از اهمیت زیادی برخوردار است، به‌طوری که یک انتخاب نادرست می‌تواند ژنوتیپ‌هایی را در یک گروه قرار دهد که شباهت زیادی با یک‌دیگر نداشته باشند. تجزیه و تحلیل معیارهای مختلف فاصله و روش‌های متفاوت تجزیه خوشه‌ای با استفاده از تجزیه تابع تشخیص در این مطالعه نشان داد که استفاده از معیار فاصله اقلیدسی بر اساس داده‌های استاندارد شده و انجام تجزیه خوشه‌ای با یکی از روش‌های حداقل واریانس "وارد" و UPGMA گروه‌بندی بهتری از ژنوتیپ‌های کلزا انجام می‌دهند. تجزیه تابع تشخیص ژنوتیپ‌های مورد مطالعه را در دو گروه قرار داد و بر اساس آن ژنوتیپ‌های Option500 و RGSS003 از سایر ژنوتیپ‌ها تفکیک شدند. نتایج حاصل از آزمون‌های T^2 هتلینگ، پلات CCC و تجزیه واریانس چند متغیره نیز گروه‌بندی حاصل از تجزیه تشخیص را مورد تأیید قرار دادند. با این حال توصیه می‌شود که در هر نوع مطالعه‌ای، روش‌های مختلف تجزیه خوشه‌ای با استفاده از آزمون‌هایی نظیر تجزیه تابع تشخیص، T^2 هتلینگ، پلات CCC و تجزیه واریانس چند متغیره مورد مقایسه و آزمون قرار گیرند تا بهترین روش در گروه‌بندی ژنوتیپ‌ها انتخاب و بر مبنای آن انتخاب ژنوتیپ‌ها برای استفاده در برنامه‌های به‌نژادی به‌طور کاملاً آگاهانه و صحیح انجام شود.

منابع مورد استفاده

۱. ابرشهر، م. ۱۳۸۷. واکنش ارقام بومی و اصلاح شده برنج به تنش خشکی. پایان‌نامه کارشناسی ارشد، دانشکده کشاورزی، دانشگاه گیلان.

۲. دانایی، م.، م.ر. احمدی و ع. گرامی. ۱۳۸۰. تجزیه کلاستر ارقام سویای ایران و به‌دست آوردن توابع محیطی مربوط به آنها. مجله علوم کشاورزی ایران ۳۲ (۲): ۲۹۳-۲۸۵.

۳. صبوری، ح.، م. نحوی، آ. ترابی و م. کاتوزی. ۱۳۸۷. طبقه‌بندی ارقام برنج ایرانی در سطوح مختلف پتانسیل اسمزی حاصل از سوربیتول براساس تجزیه خوشه‌ای و توابع تشخیص خطی فیشر. دهمین کنگره زراعت و اصلاح نباتات ایران، ۲۸-۳۰ مرداد، کرج، انجمن علوم زراعت و اصلاح نباتات ایران.

۴. صفایی چائی‌کار، ص. ۱۳۸۶. تحمل به تنش کمبود آب در ژنوتیپ‌های برنج. پایان نامه کارشناسی ارشد، دانشکده کشاورزی، دانشگاه گیلان.

5. Balocchi, L.O., J.V. Caballero and R.R. Smith. 2001. Characterization and agronomic variability of 125 ecotypes of *Bromus valdivianus* Phil, collected from Valdivia province. Agro. Sur. 29 (1): 64-77.
6. Fernandez, G.C.J. 2002. Discriminant analysis: A powerful classification technique in data mining. Proceedings of the twenty-Seventh Annual SAS Users Group International Conference. 14-17 Apr. 2002, Orlando, Florida, USA.
7. Jaynes, D.B., T.C. Kaspar, T.S. Colvin and D.E. James. 2003. Cluster analysis of spatiotemporal corn yield patterns in an Iowa field. Agron. J. 95(3): 574-586.
8. Jiang, J.H., R. Tsenkova and Y. Ozaki. 2001. Principle discriminant variate method for classification of multicollinear data: Principle and Applications. Anal. Sci. 17: 471-474.
9. Jobson, J.D. 1992. Applied Multivariate Data Analysis. Volum Π, Categorical and Multivariate Methods. Springer-Verlag, New York.
10. Mendez, M.A., C. Hodar, C. Vulpe, M. Gonzalez and V. Cambiazo. 2002. Discriminant analysis to evaluate clustering of gene expression data. Federation of Eur. Biochem. Soc. 522: 24-28.
11. Moreda, A.P., A. Fisher and S.J. Hill. 2003. The classification of tea according to region of origin using pattern recognition techniques and trace metal data. J. Food Composition and Anal. 16 (2): 195-211.
12. Romesburg, H.C. 1990. Cluster Anal.For Research. Krieger Pub., Malabar, Florida.
13. Welling, M. 2006. Fisher Linear Discriminant Anal. Univ. of Toronto Pub., Canada.

[DOR: 20.1001.1.24763594.1388.13.47.43.1]