

تحقیقات غلات

دوره هشتم / شماره اول / بهار ۱۳۹۷ (۷۲-۵۷)

 استنباط بیزی برای مطالعه کنترل ژنتیکی تحمل به تنش کم آبی در گندم با استفاده
 از روش LASSO
سیده فاطمه دانیالی^{۱*}، پرویز صفری^۲ و مهدی رحیمی^۳

تاریخ دریافت: ۱۳۹۶/۳/۲۰

تاریخ پذیرش: ۱۳۹۶/۱۲/۷

چکیده

خشکی مهم‌ترین تنش غیرزیستی است که به‌طور جدی تولید و کیفیت محصول گندم ایران را تحت تاثیر قرار می‌دهد. اطلاعات در باره کنترل ژنتیکی توارث تحمل به تنش برای تعیین نوع برنامه اصلاحی و تولید ارقام متحمل ضروری است. در این مطالعه، استنباط بیزی با به‌کارگیری روش LASSO برای شناسایی مهم‌ترین آثار ژنتیکی مرتبط با تحمل به خشکی در چارچوب روش تجزیه میانگین نسل‌ها مورد استفاده قرار گرفت. به‌همین منظور آزمایش‌هایی شامل دو جفت تلاقی بین ارقام حساس و متحمل (هامون × دریا و سپاهان × مروارید) و نسل‌های حاصل از تلاقی آن‌ها برای دو سال به‌صورت کرت‌های خردشده در دو شرایط آبیاری مطلوب و قطع آبیاری از زمان گرده‌افشانی بر پایه طرح بلوک‌های کامل تصادفی با سه تکرار اجرا شدند. استنباط بیزی یک روش جایگزین مناسب است که دانش پیشین موجود (توزیع پیشین) را با اطلاعات حاصل از داده‌ها ترکیب می‌کند و یک توزیع احتمالی (توزیع پسین) را نتیجه می‌دهد که از تمام اطلاعات برای تفسیر ساختار ژنتیکی برخوردار است. LASSO یک روش موثر برای اعمال انقباض و گزینش بر آثار مدل است. ضرایب آثار غیرمهم در مدل به سمت صفر منقبض می‌شوند و از مدل کنار گذاشته می‌شوند، در حالی که انقباض کم‌تری برای ضرایب مهم به‌کار برده می‌شود. از آنجا که عمل ژنی افزایشی، غالبیت و اپیستازی در توارث عملکرد دانه در شرایط تنش و غیرتنش درگیر بودند، روش‌هایی که از تمام آثار ژنی استفاده کنند، مانند گزینش دوره‌ای و به دنبال آن روش شجره‌ای می‌تواند روشی سودمند برای بهبود تحمل به تنش خشکی باشد.

واژه‌های کلیدی: اثر اپیستازی، اثر غالبیت، اثر افزایشی، تجزیه میانگین نسل‌ها، عمل ژن

- ۱- دانشجوی دکتری، گروه به‌نژادی و بیوتکنولوژی گیاهی، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران
- ۲- دانش‌آموخته دکتری، گروه به‌نژادی و بیوتکنولوژی گیاهی، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران
- ۳- استادیار، گروه بیوتکنولوژی، پژوهشکده علوم محیطی، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی

صنعتی و فناوری پیشرفته، کرمان، ایران

* نویسنده مسئول: danyalisf@gmail.com

ساده‌تر حاصل می‌شود. در چارچوب بی‌زی، لیکو و انتزوفراس (Lykou and Ntzoufras, 2013) یک مدل LASSO بی‌زی ارائه دادند که مزایای هر دو روش انقباض و گزینش متغیر را به کار می‌برد. انقباض از طریق استفاده از محصول توزیع‌های پیشین دو نمایی مستقل برای ضرایب مدل به دست می‌آید. در حالی که گزینش متغیر از طریق استفاده از نشانگرهای دودویی در بر گرفتن متغیر موجود در پیش‌بینی کننده خطی کسب می‌شود.

محبوبیت روش‌های آماری بی‌زی به دلیل توانایی آنها در خلاصه کردن عدم قطعیت در یک شکل احتمالی، روز به روز در علم ژنتیک بیشتر شده است (Waldmann *et al.*, 2008). تفاوت اصلی بین تئوری آمار کلاسیک و استنباط بی‌زی آن است که روش بی‌زی پارامترها را به‌عنوان متغیرهای تصادفی در نظر می‌گیرد که توسط توزیع پیشین توصیف می‌شوند. سپس روش بی‌زی داده‌های مشاهده شده را در نظر می‌گیرد و اعتقادات پیشین (توزیع پیشین) را با درست‌نمایی متداول ترکیب می‌کند تا توزیع پسین پارامتر مورد نظر که استنباط آماری بر اساس آن انجام می‌شود، به دست آید (Ntzoufras, 2011). در مرحله بعد نیز نمونه‌برداری از توزیع پسین با استفاده از روش‌های مونت کارلوی زنجیر مارکف (MCMC) انجام می‌شود که بخش "مونت کارلو" به فرآیند شبیه‌سازی تصادفی و بخش "زنجیر مارکف" به فرآیندهای نمونه‌برداری یک مقدار جدید از توزیع پسین به شرط مقدار قبلی اشاره دارد. این فرآیند تکراری، موجب تولید نمونه تصادفی از توزیع پسین می‌شود (Lynch, 2007). الگوریتم‌های نمونه‌برداری Metropolis- Hasting (Metropolis *et al.*, 1953;) و نمونه‌برداری Gibbs (Geman and Geman, 1970) و نمونه‌برداری (Geman, 1984; Gelfand and Smith, 1990) دو روش کارآمد MCMC هستند.

استنباط بی‌زی در مطالعات ژنتیک کمی در گیاهان مورد استفاده قرار گرفته است. والدمن و همکاران (Waldman *et al.*, 2008) یک الگوریتم نمونه‌برداری Gibbs برای برآورد واریانس‌های افزایشی و غالبیت در مدل مخلوط برای *scotspine* معرفی کردند. الگوریتم نمونه‌برداری MCMC دیگر در مورد مدل مخلوط توسط متو و همکاران (Mathew *et al.*, 2012) برای برآورد پارامترهای ژنتیکی داده‌های لاین‌های جو پیشنهاد شد. بالستر و همکاران (Balestre *et al.*, 2012) برای مطالعه کنترل ژنتیکی مقاومت به لکه برگ‌گی خاکستری در ذرت با استفاده از تجزیه

گندم (*Triticum aestivum* L.) مهم‌ترین گیاه زراعی در ایران است، جایی که خشکی اصلی‌ترین تنش غیرزیستی محسوب می‌شود و عملکرد و کیفیت دانه را در طی دوره پر شدن دانه به شدت تحت تاثیر قرار می‌دهد و موجب کاهش عملکرد و کیفیت دانه می‌شود (Nezhadahmadi *et al.*, 2013). انتخاب روش‌های گزینش و اصلاحی برای بهبود ژنتیکی تحمل به تنش هر نبات عمدتاً به شناخت نوع و مقدار نسبی اجزای ژنتیکی برای صفات مختلف در مواد گیاهی تحت بررسی بستگی دارد. تجزیه میانگین نسل‌ها که توسط متر و جینکز (Mather and Jinks, 1971) و کرسی و پونی (Kearsey and Pooni, 1998) پیشنهاد شد، یک روش موثر برای بررسی اهمیت آثار ژنتیکی در توارث صفات است که در آن ارزش ژنتیکی مربوط به هر نسل با یک مدل خطی شامل آثار افزایشی، غالبیت و اپیستازی دو ژنی ارزیابی می‌شوند و از آزمون مقیاس مشترک بر اساس روابط بین میانگین نسل‌ها برای تفسیر آثار ژنتیکی استفاده می‌شود. با این وجود، این روش چند عیب دارد که ممکن است کاربرد آن را محدود سازد. این روش نمی‌تواند به طور مستقیم مدل‌های خطی که در آنها تعداد پارامترها برابر یا بیشتر از تعداد نسل‌ها است را مدیریت کند. از سوی دیگر این امکان وجود دارد که آثار اصلی و اپیستازی بیش از حد یا کمتر از حد برآورد شوند چرا که این آثار در روش کم‌ترین مربعات وزنی بدون خطا برآورد می‌شوند (Balestre *et al.*, 2012).

گزینش و انقباض متغیر به‌عنوان دو روش کلی برای مدیریت مدل‌های خطی پیشنهاد شده‌اند که می‌توان از آن‌ها به‌عنوان روش‌های جایگزین برای شناسایی و برآورد مهم‌ترین آثار مدل در تجزیه میانگین نسل‌ها استفاده کرد. گزینش متغیر یک تکنیک مهم برای کاهش متغیرهای مدل و شناسایی متغیرهای موثر بر مدل است. به جای حذف تمام متغیرهای غیرمعنی‌دار از مدل، در تجزیه انقباض، تمام متغیرهای مدل در نظر گرفته می‌شوند، ولی برآوردها برای متغیرهای غیرمعنی‌دار به سمت صفر منقبض می‌شوند (Xu, 2007). تیبشیرانی (Tibshirani, 1996) روش LASSO را برای انقباض و گزینش در مسائل رگرسیونی پیشنهاد داد. مدل LASSO، مجموع مربعات باقی‌مانده و ناگزیر مجموع مقادیر مطلق ضرایب مدل را به حداقل می‌رساند. این محدودیت خاص باعث می‌شود تا برخی از ضرایب مدل دقیقاً صفر باشند و از این‌رو مدل‌هایی با تفسیر

آبیاری در کرت‌های اصلی شامل دو سطح (آبیاری و قطع آبیاری) و نسل‌ها در کرت‌های فرعی با شش و نه سطح (به ترتیب برای سال اول و دوم) قرار گرفتند. جهت تهیه بستر کشت، شخم نسبتاً عمیقی در اواسط فروردین زده شد و سپس عملیات دیسک‌زنی اجرا شد. کرت‌های آزمایشی در سال اول (۱۳۹۴) به ابعاد $۳ \times ۶/۳$ متر و در سال دوم (۱۳۹۵) به ابعاد $۳ \times ۸/۳$ متر بود و فاصله بین تکرارها در هر دو سال $۱/۵$ متر بود. در وسط هر کرت یک راهرو برای ایجاد سهولت رفت و آمد و اندازه‌گیری در نظر گرفته شد و به این ترتیب هر کرت به دو قسمت برای کشت تقسیم شد. طول هر ردیف کشت ۹۰ سانتی‌متر بود که در آن ۱۷ بذر با فاصله ۵ سانتی‌متر کشت شدند. فاصله بین ردیف‌ها ۱۵ سانتی‌متر در نظر گرفته شد و در ابتدا و انتهای ردیف‌های کشت، ردیف‌هایی نیز به‌عنوان حاشیه کشت شدند. برای تلاقی دوم در سال اول ردیف‌های خالی کرت‌ها با ارقام والدی پر شدند.

واحدهای آزمایشی تعداد ردیف‌های متفاوتی داشتند که بستگی به یکنواختی هر نسل داشت. برای نسل‌های بدون تفرق P_1 و P_2 دو ردیف کشت شدند و برای نسل F_2 در سال اول برای تلاقی اول پنج ردیف (۷۵ بوته در هر کرت اصلی) و برای تلاقی دوم دو ردیف (۳۰ بوته در هر کرت اصلی) و در سال دوم برای تلاقی اول چهار ردیف (۶۰ بوته در هر کرت اصلی) و برای تلاقی دوم پنج ردیف (۷۵ بوته در هر کرت اصلی) کشت شدند. تعداد خانواده‌های مورد استفاده برای نسل‌های مختلف در جدول ۱ ارائه شده است. عملیات کاشت در اواسط فروردین به‌صورت دستی انجام شد. پس از کشت تا زمان گرده‌افشانی، آبیاری در هر دو شرایط تیمار آبیاری و قطع آبیاری به‌طور یکسان انجام شد. مقدار نیتروژن مورد نیاز گیاه از منبع اوره به مقدار ۲۰۰ کیلوگرم در هکتار استفاده شد. پخش کود همراه با آب آبیاری در طی دو مرحله انجام گرفت. وجین علف‌های هرز به دفعات مورد نیاز به صورت دستی انجام گرفت و برداشت محصول در اوایل شهریور به صورت دستی انجام شد.

مواد گیاهی مورد استفاده در این آزمایش، نسل‌های مختلف حاصل از دو جفت تلاقی ارقام گندم بهاره هامون \times دریا (C1) و سپاهان \times مروارید (C2) بود. بذر ارقام والدی از بانک بذر موسسه تحقیقات اصلاح و تهیه نهال و بذر تهیه شدند. دریا و مروارید، حساس و والدین هامون و سپاهان متحمل به شوری و خشکی هستند.

استنباط بی‌زی برای مطالعه ژنتیکی تحمل به کم‌آبی در گندم

میانگین نسل‌ها از استنباط بی‌زی بر اساس روش انقباض پیشنهادی زو (Xu, 2003) استفاده کردند و کنترل ژنتیکی مقاومت به این بیماری را به اثرات افزایشی نسبت دادند و روش یاد شده را برای مدیریت مدل‌های پیچیده، روشی ممتاز معرفی کردند چرا که توانایی تنظیم اثرات افزایشی، غالبیت، اپیستازی و واریانس آنها را در یک مدل ساده دارا است. وانگ و همکاران (Wang et al., 2011) روش LASSO مختلط تطبیقی را برای شناسایی مکان‌های ژنی کنترل‌کننده صفات کمی و اثرات اپیستازی در جمعیت‌های گیاهی پیشنهاد دادند. آنها نشان دادند که در مقایسه با روش LASSO اصلی و LASSO تطبیقی این روش به دلیل این که وزن‌های مختلف را برای به کارگیری مقدار انقباض به هر یک از پارامترها اختصاص می‌دهد و همچنین اثرات تصادفی را نیز در مدل در نظر می‌گیرد، بنابراین عملکرد مدل را افزایش می‌دهد. وانگ و همکاران (Wang et al., 2012) برای پیش‌بینی ارزش‌های ژنتیکی صفات کمی برخوردار از اثرات اپیستازی در جمعیت‌های گیاهی از روش LASSO مختلط تطبیقی استفاده کردند و ۲۸۰ توده را در برنامه اصلاح گندم نبراسکا مورد بررسی قرار دادند. نتایج نشان داد که LASSO مختلط تطبیقی را می‌توان به طور موفقیت آمیزی در پیش‌بینی ارزش‌های ژنتیکی با ترکیب اثرهای اصلی و اپیستازی نشانگرها به کار برد. دقت پیش‌بینی با گنجاندن اثرات اپیستازی دو مکان ژنی به طور فزاینده‌ای افزایش یافت.

هدف از این مطالعه استنباط بی‌زی اثرات ژنتیکی برای عملکرد دانه گندم با استفاده از نسل‌های مختلف در شرایط مطلوب و تنش بود. در این پژوهش از روش LASSO بی‌زی برای برآورد و شناسایی مهم‌ترین اثرات ژنتیکی مدل‌ها استفاده شد.

مواد و روش‌ها

طرح و مواد گیاهی

نسل‌های F_2 ، F_3 ، BC_1S_1 و BC_2S_1 (در سال اول و دوم) و F_4 ، BC_1S_2 و BC_2S_2 (در سال دوم) به‌همراه والدین برای انجام آزمایش در مزرعه پژوهشی دانشگاه تحصیلات تکمیلی صنعتی و فناوری‌های پیشرفته کرمان کشت شدند. آزمایش‌ها به‌صورت کرت‌های خردشده تحت دو شرایط آبیاری مطلوب و قطع آبیاری از زمان گرده‌افشانی بر پایه بلوک‌های کامل تصادفی با سه تکرار اجرا شدند. تیمار

جدول ۱- تعداد خانواده‌های ارزیابی شده در نسل‌های مختلف برای تلاقی‌های مختلف
Table 1. The number of evaluated families of different generations crosses

Generations	C1 (Hamoon×Darya)		C2 (Sepahan×Morvarid)	
	2015	2016	2015	2016
F ₃	30	20	17	20
F ₄	-	20	-	15
BC ₁ S ₁	20	13	16	15
BC ₂ S ₁	20	12	15	14
BC ₁ S ₂	-	16	-	15
BC ₂ S ₂	-	16	-	15

که در آن Y_i مقدار مشاهده شده میانگین نسل‌ها، x_{i1} تا x_{i5} به ترتیب ضرایب مربوط به اثر افزایشی (a)، اثر غالبیت (d)، آثار اپیستازی افزایشی × افزایشی (aa)، افزایشی × غالبیت (ad) و غالبیت × غالبیت (dd) هستند. β_0 میانگین کل، $\beta = (\beta_1, \dots, \beta_5)^T$ بردار ضرایب مدل منعکس‌کننده آثار [a]، [d]، [aa]، [ad] و [dd] مرتبط با نسل‌ها هستند و e_i باقی‌مانده مدل با توزیع $N(0, \sigma^2)$ است.

LASSO دارای یک تفسیر بیزی ساده است چرا که برآوردهای آن را می‌توان زمانی که توزیع‌های پیشین دو نمایی (DE) مستقل برای β مورد استفاده قرار می‌گیرد، به صورت حالت توزیع پسین به دست آورد. با معرفی یک بردار از نشانگرهای γ دودویی که مشخص می‌کند کدام متغیرها در مدل باقی بمانند (با $\gamma_i = 1$) یا خیر (با $\gamma_i = 0$) می‌توان به فرآیند گزینش متغیر دست یافت (Lykou and Ntzoufras, 2013). برای تنظیم مدل بیزی، درست‌نمایی معمول مدل در ترکیب با نشانگرهای دودویی در بر گرفتن متغیر در مدل $\gamma = (\gamma_1, \dots, \gamma_p)$ مشابه روش‌های پیشنهادی کائو و مالیک (Kuo and Mallick, 1998) و دلاپورتاس و همکاران (Dellaportaset al., 2002) در نظر گرفته شد. سپس یک مجموعه از توزیع‌های پیشین دو نمایی برای هر یک از ضرایب مدل β_j به منظور به کارگیری یک نوع انقباض LASSO درون هر مدل فرض شد. از این رو مدل را می‌توان توسط روابط زیر خلاصه کرد (Lykou and Ntzoufras, 2013):

$$\mathbf{Y} | \beta, \tau, \gamma \sim N(\mathbf{X}\mathbf{D}_\gamma \beta, \tau^{-1} I_n) \quad (2)$$

$$\mathbf{D}_\gamma = \text{diag}(\gamma_1, \dots, \gamma_p) \quad \text{که:} \quad (3)$$

$$\beta_j | \tau \sim DE\left(0, \frac{1}{\tau \lambda}\right), \quad \text{برای } j = 1, \dots, p \quad (4)$$

$$\gamma_j \sim \text{Bernoulli}(\pi_j) \quad (5)$$

برای اجرای صحیح نمونه‌برداری و حذف اثرات حاشیه، بوته‌های واقع در ردیف‌های ابتدایی و انتهایی هر کرت و نیز بوته‌های ابتدا و انتهای هر ردیف کشت در نظر گرفته نشدند و اندازه‌گیری‌ها برای عملکرد دانه (گرم در بوته) با برداشت ۱۰ بوته از هر ردیف انجام گرفت. پیش از انجام تجزیه‌ها، آزمون نرمال بودن توزیع باقیمانده‌ها برای صفت اندازه‌گیری شده انجام گرفت. برای انجام تجزیه مرکب آزمایش‌ها، تصادفی کردن طرح‌های پایه در سال‌های مختلف به صورت جداگانه انجام شد و برای اطمینان از یکنواختی یا همگنی اشتباه‌های آزمایشی از آزمون بارتلت استفاده شد.

تجزیه واریانس مرکب (Gomez and Gomez, 1984) با استفاده از برنامه آماری SAS ver. 9.0 (SAS, 2002) انجام شد و سپس با توجه به معنی‌داری برهمکنش نسل × سال، تجزیه واریانس به‌طور جداگانه برای هر سال برای شناسایی تفاوت‌های معنی‌دار میان نسل‌ها انجام گرفت. با توجه به معنی‌داری برهمکنش نسل × تیمار، تجزیه‌های بعدی به‌طور جداگانه در هر سطح تنش انجام شد. تجزیه میانگین نسل‌ها به روش متداول بر اساس روش کمترین مربعات توصیف شده توسط متر و جینکز (Mather and Jinks, 1982) به کمک نرم‌افزاری که برای این منظور طراحی شد، انجام گرفت. این نرم‌افزار با زبان برنامه‌نویسی ویژوال استدیو C# (سی‌شارپ) به کمک فایل‌های کتابخانه‌ای dll و بانک اطلاعاتی Access طراحی شد.

استنباط بیزی

مدل خطی زیر برای میانگین هر نسل در نظر گرفته

شد:

$$Y_i = \beta_0 + \sum_{j=1}^{p=5} \beta_j x_{ij} + e_i \quad i = 1, \dots, n \quad (1)$$

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_5 x_{i5} + e_i$$

$$\tau \sim \text{Gamma}(c, d) \quad (5)$$

در این روابط، $\tau = 1/\sigma^2$ دقت مدل، λ پارامتر انقباض، $\text{Bernouli}(\pi_j)$ توزیع برنولی با احتمال موفقیت π و $\text{Gamma}(c, d)$ توزیع گاما با میانگین c/d و واریانس c/d^2 است. مقادیر مختلف λ با سطوح مختلف انقباض مطابقت دارد. مقادیر کوچک λ مطابق است با واریانس بزرگ که معرف اطلاعات کم در مدل بیزی است و اساساً هیچ انقباضی بر پارامترهای مدل اعمال نمی‌کند. از سوی دیگر، مقادیر بزرگ λ مطابق است با توزیع پیشین قوی که در آن β به علت سطوح بالای انقباض به صفر نزدیک می‌شود (Lykou and Ntzoufras, 2011). استدلال‌های انتخاب مناسب مقدار λ توسط لیکو و انتزوفراس (Lykou and Ntzoufras, 2011) ارایه شده است. برای اطمینان از این که λ معنی یکسانی بر تمام آثار مدل دارد، بدون از دست‌دادن کلیت فرض می‌شود که تمام متغیرهای توضیحی دارای میانگین صفر و واریانس معادل یک هستند. در چارچوب بیزی، متغیرهای پاسخ و آثار استاندارد می‌شوند. بنابراین، پارامتر ثابت β_0 (m در تجزیه میانگین نسل‌ها) در مدل در نظر گرفته نمی‌شود، چرا که به‌طور طبیعی مقدار آن معادل صفر است. در تنظیم توزیع پیشین، واریانس σ^2 توسط پارامتر دقت متناظر τ به‌منظور سازگاری با نمادگذاری نرم‌افزار WinBUGS جایگزین شده است. برای τ از مقادیر پیشین کم استفاده شد. به‌این ترتیب میانگین پیشین آن برابر یک تنظیم و واریانس پیشین آن مقداری بزرگ در نظر گرفته شد. در این مطالعه، $c = d = 0.0001$ استفاده شد. برای توزیع برنولی، احتمال موفقیت $\pi = 0.5$ در نظر گرفته و با توجه به استدلال‌های انتخاب مقدار مناسب λ ، تعداد پارامترهای مدل و تعداد مشاهدات، $\lambda = 0.06$ برای انجام تجزیه انتخاب شد.

برای نمونه‌برداری از توزیع پسین از الگوریتم نمونه‌برداری Gibbs استفاده شد. توزیع پسین شرطی β_j با توزیع پیشین برای $\gamma_j = 0$ منطبق است، در حالی که وقتی $\gamma_j = 1$ است، این توزیع مخلوطی از توزیع‌های نرمال برش داده شده است، یعنی:

$$\beta_j | y, \tau, \beta_j, \gamma_j, \gamma_j = 0 \sim \text{DE}\left(0, \frac{1}{\tau\lambda}\right) \quad (6)$$

$$\beta_j | y, \tau, \beta_j, \gamma_j, \gamma_j = 1,$$

$$\Omega_j \sim \Omega_j \text{TN}(\mu_j^-, s_j^2, \beta_j < 0) + (1 - \Omega_j) \text{TN}(\mu_j^+, s_j^2, \beta_j \geq 0) \quad (7)$$

که β_j و γ_j به ترتیب بردارهای β و γ بدون β_j و γ_j هستند، $I(A)$ تابع نشانگر است که وقتی صحیح است، یک و در غیر این صورت صفر در نظر گرفته می‌شود و $\text{TN}(\mu, \sigma^2, A)$ توزیع نرمال برش داده شده در زیر مجموعه $A \subset \mathbb{R}$ با تابع چگالی زیر است:

$$f_{\text{TN}}(x; \mu, \sigma^2, A) = \frac{f_N(x; \mu, \sigma^2)}{\int_A f_N(x; \mu, \sigma^2) dx} I(x \in A) \quad (8)$$

که $f_N(x; \mu, \sigma^2)$ چگالی یک توزیع نرمال با میانگین μ و واریانس σ^2 ارزیابی شده در x است. در نتیجه چگالی‌های توزیع‌های نرمال برش داده شده موجود در رابطه (7) به ترتیب توسط روابط زیر به دست می‌آیند:

$$f_{\text{TN}}(\beta_j; \mu_j^-, s_j^2, \beta_j < 0) = \frac{f_N(\beta_j; \mu_j^-, s_j^2)}{\Phi(-\mu_j^-/s_j)} I(\beta_j < 0) \quad (9)$$

و

$$f_{\text{TN}}(\beta_j; \mu_j^+, s_j^2, \beta_j \geq 0) = \frac{f_N(\beta_j; \mu_j^+, s_j^2)}{\Phi(\mu_j^+/s_j)} I(\beta_j \geq 0) \quad (10)$$

که $\Phi(x)$ ، تابع چگالی احتمال توزیع نرمال استاندارد شده است. میانگین‌ها و واریانس توزیع‌های نرمال برش داده شده توسط روابط زیر محاسبه می‌شوند:

$$\mu_j^- = \frac{c_j + \lambda}{\|X_j\|^2} \quad \text{و} \quad \mu_j^+ = \frac{c_j - \lambda}{\|X_j\|^2}, \quad (11)$$

$$c_j = X_j^T (e + \beta_j X_j) \quad \text{و} \quad s_j^2 = \frac{1}{\tau \|X_j\|^2}$$

که X_j زامین ستون ماتریس X ، $e = y - X\beta$ بردار

مقادیر باقی‌مانده‌ها است، اما $\|K\|^2 = \sum_{i=1}^n K_i^2$ است.

افزون بر این، Ω_j یک بردار پارامتر دودویی تعیین کننده علامت β_j است. احتمال پسین شرطی کامل $\Omega_j = 1$ توسط رابطه زیر به دست می‌آید:

$$\begin{aligned} w_j &= \text{P}(\Omega_j = 1 | y, \tau, \beta_j, \gamma_j, \gamma_j = 1) \\ &= \text{P}(\beta_j < 0 | y, \tau, \beta_j, \gamma_j, \gamma_j = 1) \\ &= \frac{\Phi(-\mu_j^-/s_j) / f_N(0, \mu_j^-, s_j^2)}{\Phi(-\mu_j^-/s_j) / f_N(0, \mu_j^-, s_j^2) + \Phi(\mu_j^+/s_j) / f_N(0, \mu_j^+, s_j^2)} \end{aligned} \quad (12)$$

زمانی که $\gamma_j = 1$ است:

* Ω_j از یک توزیع برنولی با احتمال موفقیت w_j ارایه شده در رابطه (12) تولید می‌شود.

پارامترهای نمونه برداری شده از ۱۰۰۰ تکرار اول از تجزیه حذف شدند (دوره گرم کردن). به دلیل ویژگی مارکف، نمونه‌های کسب شده از زنجیر مارکف مستقل نیستند. بنابراین، برای کاهش همبستگی سریالی، مشاهدات در هر ۱۰۰ تکرار ذخیره شدند. برای تولید توزیع پسین مدل و نمونه برداری از آن با استفاده از الگوریتم نمونه برداری Gibbs از نرم افزار WinBUGS (Spiegelhalter et al.,) WinBUGS (2003) استفاده شد. WinBUGS یک بسته نرم افزاری رایگان است که نمونه برداری Gibbs را انجام می‌دهد (www.mrc-bsu.com.ac.uk/bugs).

نتایج و بحث

نتایج تجزیه واریانس مرکب کرت‌های خرد شده در دو سال با ثابت در نظر گرفتن اثر نسل و تصادفی در نظر گرفتن اثر سال در جدول ۲ ارائه شده است. اثر سال برای هر دو تلاقی معنی دار بود که نشان دهنده وجود اختلاف بین صفت در سال‌های مختلف بود. اثر تیمار در هر دو تلاقی معنی دار بود و عملکرد دانه را تحت تاثیر قرار داد. اثر نسل برای هر دو تلاقی معنی دار بود و نشان داد که بین نسل‌ها در هر دو تلاقی تفاوت وجود دارد. معنی دار شدن برهمکنش نسل × سال نشان داد که نسل‌های مورد بررسی در سال‌های مختلف واکنش یکسانی به شرایط آب و هوایی نشان ندادند. معنی داری برهمکنش نسل × تیمار برای عملکرد دانه در هر دو تلاقی نشان داد که در مجموع سال‌ها، نسل‌های مورد بررسی در تیمارهای مختلف واکنش متفاوتی نشان دادند.

* β_j از روابط زیر ایجاد می‌شود:

$$\begin{cases} TN(\mu_j^-, s_j^2, \beta_j < 0) & \text{اگر } \Omega_j = 1, \\ TN(\mu_j^+, s_j^2, \beta_j \geq 0) & \text{اگر } \Omega_j = 0. \end{cases} \quad (13)$$

توزیع‌های پسین شرطی کامل برای پارامترهای باقیمانده به صورت زیر هستند:

$$\tau | \beta, \gamma, y \sim \text{Gamma} \left(\frac{n}{2} + p + c, \frac{\|y - \mathbf{X}\beta\|^2}{2} + \lambda \|\beta\| + d \right) \quad (14)$$

$$\gamma_j | \beta, \tau, \gamma_j, y \sim \text{Bernoulli} \left(\frac{O_j}{1 + O_j} \right) \quad (15)$$

با:

$$\begin{aligned} O_j &= \frac{f(\gamma_j = 1 | \gamma_j, \beta, \tau^2, y)}{f(\gamma_j = 0 | \gamma_j, \beta, \tau^2, y)} \\ &= \frac{f(y | \beta, \tau, \gamma_j, \gamma_j = 1) \pi(\gamma_j, \gamma_j = 1)}{f(y | \beta, \tau, \gamma_j, \gamma_j = 0) \pi(\gamma_j, \gamma_j = 0)} \end{aligned} \quad (16)$$

روابط ذکر شده توسط لیکو و انتزوفراس (Lykou and Ntzoufras, 2013) توصیف شده‌اند.

برای نمونه برداری از توزیع پسین و برآورد پارامترهای مدل از روش‌های مونت کارلوی زنجیر مارکف (MCMC) استفاده می‌شود. در این مطالعه، نمونه برداری توسط الگوریتم نمونه بردار Gibbs (Geman and Geman,) (1984) انجام شد. نمونه برداری Gibbs شامل فرآیند تکراری کسب نمونه‌ها از توزیع پسین برای هر پارامتر و تکرار این فرآیند به روز شونده است. طول زنجیر MCMC به کار برده شده در این تحقیق، ۸۰۰۰ تکرار بود. مقادیر

جدول ۲- تجزیه واریانس مرکب عملکرد دانه (گرم در بوته) برای دو تلاقی مطالعه شده در این تحقیق

Table 1. Combined analysis of variance for grain yield (g.plant⁻¹) for both crosses studied in this research

Source of variations	df	Mean squares	
		Cross 1	Cross 2
Year	1	1.58**	12.33**
Replication (Year)	4	0.59**	0.13**
Treatment	1	1.70**	1.29**
Year × Treatment	1	0.14 ^{ns}	0.07 ^{ns}
Treatment × Replication (Year)	4	0.26**	0.06*
Generation	8	4.39**	1.43**
Treatment × Generation	8	0.18**	0.07**
Year × Generation	5	0.21**	0.70**
Year × Treatment × Generation	5	0.06 ^{ns}	0.02 ^{ns}
CV (%)		11.73	9.68

^{ns}, * and **: Not-significant and significant at 5% and 1% probability levels, respectively.

جدول ۳- تجزیه واریانس عملکرد دانه (گرم/بوته) برای سال‌های مختلف
Table 3. Analysis of variance for grain yield (g.plant⁻¹) in different years

Source of variations	2015			2016		
	df	Mean squares		df	Mean squares	
		Cross 1	Cross 2		Cross 1	Cross 2
Replication	2	1.17 ^{ns}	0.27 ^{ns}	2	0.02 ^{ns}	0.00 ^{ns}
Treatment	1	0.32 ^{ns}	1.08 ^{ns}	1	1.88 ^{ns}	0.42*
Main-plot error	2	0.16	0.11	2	0.35	0.007
Generation	5	1.45**	2.61**	8	3.57**	0.24**
Generation × Treatment	5	0.55**	0.18**	8	0.07**	0.03**
Sub-plot error	20	0.13	0.04	32	0.02	0.01
CV (%)		19.31	10.30		6.50	7.88

^{ns}, * and **: Not-significant and significant at 5% and 1% probability levels, respectively.

مونت کارلو یا خطای MC تغییرات میانگین پارامتر مورد نظر ناشی از شبیه‌سازی را نشان می‌دهد. اگر خطای MC در مقایسه با انحراف استاندارد پسین برآورد شده متناظر کم باشد، نشان‌دهنده این واقعیت است که میانگین پسین با دقت زیاد برآورد شده است (Ntzoufras, 2011). با توجه به جدول‌های ۶ و ۷ مشاهده می‌شود که خطای MC برای تمام پارامترها کم است و بنابراین میانگین‌های پسین با دقت زیاد برآورد شده‌اند.

برآورد‌های آثار ژنی برای بهترین مدل‌های برازش یافته بر اساس روش متداول در جدول‌های ۴ و ۵ ارایه شده‌اند. خلاصه‌های پسین پس از اعمال انقباض و احتمالات پسین در بر گرفتن آثار در مدل در جدول‌های ۶ و ۷ ارایه شده است. شکل‌های ۱ تا ۸ چگالی‌های توزیع پسین آثار را نشان می‌دهند. برآورد‌های بیزی آثار مهم موجود در مدل‌ها نیز در جدول‌های ۶ و ۷ ارایه شده است. خلاصه‌های پسین شامل برآورد‌هایی از میانگین پسین، انحراف استاندارد و خطای MC برای پارامترهای مورد نظر هستند. خطای

جدول ۴- برآورد پارامترهای مدل عملکرد دانه و آزمون مقیاس مشترک برای بهترین مدل برازش یافته در تلاقی اول

Table 4. Estimating model parameters of grain yield and joint scalling test of the best fitted models in cross-1

Model parameters	2015; Normal	2015; Stress	2016; Normal	2016; Stress
[m]	0.845 ± 0.333	0.744 ± 0.294	2.894 ± 0.124	2.548 ± 0.160
[a]	-1.015 ± 0.050	-0.553 ± 0.048	-1.256 ± 0.049	-1.023 ± 0.049
[d]	2.873 ± 1.098	3.123 ± 0.934	-1.145 ± 0.458	-2.042 ± 0.929
[aa]	1.176 ± 0.336	0.826 ± 0.298	-0.603 ± 0.133	-0.518 ± 0.166
[ad]	1.651 ± 0.775	-	-3.002 ± 0.400	-3.510 ± 0.418
[dd]	-	-	-	3.646 ± 1.434
X ² (df)	2.277 ^{ns} (1)	2.321 ^{ns} (2)	46.515** (4)	42.718** (3)

^{ns} and **: Not-significant and significant at 1% probability level, respectively.

جدول ۵- برآورد پارامترهای مدل عملکرد دانه و آزمون مقیاس مشترک برای بهترین مدل برازش یافته در تلاقی دوم

Table 5. estimating model parameters of grain yield and joint scalling test of the best fitted models in cross-2

Model parameters	2015; Normal	2015; Stress	2016; Normal	2016; Stress
[m]	-0.503 ± 0.306	0.744 ± 0.130	1.064 ± 0.058	1.155 ± 0.016
[a]	-0.895 ± 0.046	-0.565 ± 0.050	-0.306 ± 0.026	-
[d]	13.481 ± 1.683	3.911 ± 0.411	0.704 ± 0.190	-
[aa]	2.358 ± 0.302	0.884 ± 0.138	0.345 ± 0.069	-
[ad]	-	-1.236 ± 0.436	-	-1.561 ± 0.221
[dd]	-15.308 ± 2.426	-	-	-
X ² (df)	5.474* (1)	41.584** (1)	50.095** (5)	44.809** (7)

* and **: Significant at 5% and 1% probability levels, respectively.

تلاقی هامون × دریا (C1)

برای عملکرد دانه در سال اول تحت شرایط نرمال مشاهده می‌شود که میانگین پسین آثار [ad] و [dd] به سمت صفر منقبض شده‌اند و میانگین پسین سایر آثار دور از صفر قرار دارند. با توجه به نتایج احتمالات پسین در بر گرفتن آثار، مشاهده می‌شود که آثار [ad] و [dd] احتمالات پسین در بر گرفتن بسیار کمی داشتند (به ترتیب ۱ و ۱۱ درصد) و باید از مدل کنار گذاشته شوند. از سوی دیگر آثار [a]، [d] و [aa] به ترتیب با دارا بودن احتمالات پسین در بر گرفتن ۱۰۰، ۵۸ و ۹۳ درصد به عنوان آثار مهم باید در مدل گنجانده شوند (جدول ۶).

در روش LASSO بیزی توزیع‌های پسین مرتبط با آثار غیرمهم، پیک بلندی در صفر نشان می‌دهند، در حالی که توزیع‌های پسین مهم‌ترین آثار دور از صفر قرار می‌گیرند (Lykou and Ntzoufras, 2011). با توجه به شکل ۱ ملاحظه می‌شود که چگالی‌های توزیع پسین آثار غیرمهم مدل ([ad] و [dd]) پیک بلندی در صفر دارند، در حالی که توزیع سایر آثار دور از صفر قرار گرفته‌اند که نشان‌دهنده اهمیت زیاد آن‌ها در مدل است. توزیع‌های پسین مربوط به آثار [d] و [aa] دو مدی شامل یک مد در صفر و یک مد در یک عدد مثبت بودند (شکل ۱). با این وجود، احتمالات پسین در بر گرفتن متناظر با این دو اثر نشان‌دهنده اهمیت این آثار هستند و باید در مدل گنجانده شوند (جدول ۶ و شکل ۱). بنابراین برای عملکرد دانه در سال اول تحت شرایط نرمال مدل چهار پارامتری [aa] [d] [a] [m] کفایت می‌کند که برآوردهای بیزی مربوط به این آثار در جدول ۶ ارائه شده است. تحت شرایط تنش نیز آثار [ad] و [dd] به سمت صفر منقبض شدند و احتمالات پسین در بر گرفتن آنها نیز به ترتیب مقادیر ناچیز ۰/۰۳ و ۰/۰۴ را دارا بودند (جدول ۶ و شکل ۲). چگالی‌های توزیع پسین متناظر با این دو اثر دارای پیک بلند در صفر بودند (شکل ۲) و بنابراین باید از مدل کنار گذاشته شوند. توزیع پسین آثار [a]، [d] و [aa] به طور کامل دور از صفر قرار داشتند که نشان‌دهنده اهمیت بسیار زیاد آن‌ها در مدل است (شکل ۲). این آثار با احتمالات پسین در بر گرفتن بالای ۹۷ درصد در مدل گنجانده شدند (جدول ۶ و شکل ۲).

در سال دوم و تحت شرایط نرمال، آثار [a]، [d] و [ad] مهم‌ترین آثار مدل تشخیص داده شدند. توزیع‌های پسین

این آثار دور از صفر قرار داشتند و احتمالات پسین در بر گرفتن آن‌ها در مدل به ترتیب ۱۰۰، ۱۰۰ و ۸۳ درصد بودند (جدول ۶ و شکل ۳). اثر [dd] با احتمال پسین در بر گرفتن ۳۲ درصد اهمیت کم‌تری در مدل داشت و به همراه اثر [aa] که به شدت به سمت صفر منقبض شده بود و توزیع پسین آن روی صفر قرار داشت، از مدل کنار گذاشته شدند (جدول ۶ و شکل ۳). تحت شرایط تنش، آثار [d] و [dd] بیش‌ترین انقباض را به سمت صفر داشتند و با دارا بودن احتمالات پسین در بر گرفتن یک درصد از مدل حذف شدند (جدول ۶ و شکل ۴). اثر [a] با دارا بودن احتمال پسین در بر گرفتن ۱۰۰ درصد به همراه آثار [ad] و [aa] با احتمالات پسین در بر گرفتن در مدل ۶۶ و ۷۸ درصد در مدل حفظ شدند (جدول ۶ و شکل ۴).

تلاقی سپاهان × مروارید (C2)

نتایج تجزیه LASSO بیزی برای عملکرد دانه در سال اول تحت شرایط نرمال نشان داد که اثر [ad] بیش‌ترین انقباض را به سمت صفر داشت و با دارا بودن احتمال پسین در بر گرفتن ۲۳ درصد از مدل کنار گذاشته شد (جدول ۷ و شکل ۵). آثار [a]، [d]، [aa] و [dd] با دارا بودن احتمالات پسین در بر گرفتن ۱۰۰ درصد مهم‌ترین آثار مدل بودند (جدول ۷ و شکل ۵). نتایج تحت شرایط تنش نشان داد که تمام آثار باید در مدل حفظ شوند. توزیع پسین این آثار کاملاً دور از صفر قرار داشت و احتمالات پسین در بر گرفتن آن‌ها در مدل بالای ۹۸ درصد بود (جدول ۷ و شکل ۶).

در سال دوم تحت شرایط نرمال، میانگین پسین آثار [d]، [ad] و [dd] به شدت به سمت صفر منقبض شدند و با دارا بودن احتمالات پسین بسیار ناچیز در بر گرفتن در مدل از مدل کنار گذاشته شدند (جدول ۷ و شکل ۷). آثار [a] و [aa] با دارا بودن احتمالات پسین در بر گرفتن ۱۰۰ درصد در مدل قرار داده شدند. چگالی‌های توزیع پسین مربوط به این آثار به طور کامل دور از صفر قرار داشتند (جدول ۷ و شکل ۷). تحت شرایط تنش فقط اثر اپیستازی افزایشی × غالبیت در مدل گنجانده شد و باقی آثار به شدت به سمت صفر منقبض شدند (جدول ۷ و شکل ۸). توزیع‌های پسین مربوط به آثار [a]، [d]، [aa] و [dd] کاملاً روی صفر قرار داشتند (شکل ۸).

جدول ۶- خلاصه‌های پسین پس از اعمال انقباض، احتمالات پسین در بر گرفتن در مدل عملکرد دانه و برآوردهای بیزی آثار مدل در تلاقی اول

Table 6. Posterior summaries after applying shrinkage, posterior inclusion probabilities and Bayesian estimations of grain yield model effects in cross-1

Grain yield	Parameter	Shrinkage			Posterior inclusion probabilities	Bayesian estimation
		Mean	Standard deviation	MC error		
First year (Normal)	m	-	-	-	-	1.43
	[a]	-0.390	0.061	0.001	1.0	-0.791
	[d]	0.156	0.171	0.014	0.583	-2.380
	[aa]	0.287	0.119	0.008	0.936	0.180
	[ad]	0.001	0.023	8.2E-4	0.019	-
	[dd]	-0.004	0.101	0.008	0.118	-
First year (Stress)	m	-	-	-	-	0.735
	[a]	-0.193	0.052	0.003	0.976	-0.359
	[d]	0.327	0.073	0.003	1.0	3.146
	[aa]	0.294	0.054	0.001	1.0	1.141
	[ad]	-0.003	0.027	0.002	0.030	-
	[dd]	-0.009	0.048	0.003	0.045	-
Second year (Normal)	m	-	-	-	-	2.956
	[a]	-0.616	0.099	0.007	1.0	-1.280
	[d]	-0.273	0.137	0.012	1.0	-2.037
	[aa]	-6.9E-5	0.009	3.0E-5	0.005	-
	[ad]	-0.191	0.105	0.008	0.835	-2.873
	[dd]	0.086	0.140	0.13	0.322	-
Second year (Stress)	m	-	-	-	-	2.125
	[a]	-0.608	0.114	0.009	1.0	-1.148
	[d]	-1.89E-4	0.005	1.10E-4	0.012	-
	[aa]	0.078	0.062	0.004	0.667	1.206
	[ad]	-0.193	0.122	0.010	0.782	-2.713
	[dd]	3.69E-4	0.005	1.09E-4	0.011	-

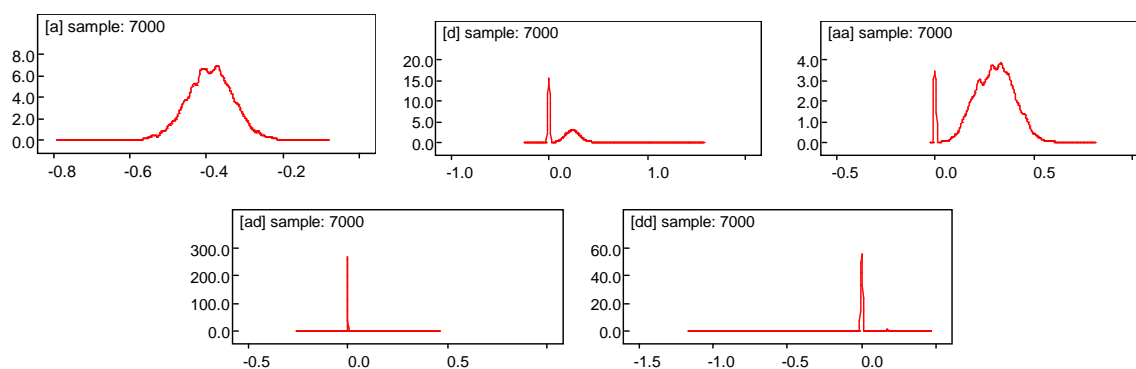
زمان‌های نزدیک یکدیگر بسیار همبسته هستند (ویژگی مارکف). با این وجود، مقادیر زمان‌های دور از هم تقریباً مستقل هستند (Ntzoufras, 2011). از میان مزایای از قبل شناخته شده تجزیه بیزی نسبت به تجزیه متداول، استفاده بالقوه از توزیع‌های پیشین، انعطاف در مدل‌های پیچیده، فواصل اطمینان دقیق و برآوردهای خطای کم‌تر را می‌توان ذکر کرد (Balestreet *et al.*, 2012). افزون بر این، استنباط‌های مستقیم از توزیع احتمال پسین و فواصل اطمینان دقیق، کاربرد موثر این روش را در استنباط‌های پارامترهای ژنتیکی امکان‌پذیر ساخته است. با استفاده از این روش امکان به‌روز رسانی دانش اصلاح‌گر در باره ژنتیک کنترل صفت بر اساس نتایج خود وجود دارد و از توزیع‌های پیشین مناسب می‌توان برای آثار اصلی و اپیستازی در مطالعات آتی استفاده کرد (Balestreet *et al.*, 2012).

در رویکرد بیزی، آنچه در باره پارامتر مشخص است (معرف توزیع پیشین) با اطلاعات حاصل از داده‌ها برای کسب توزیع پسین ترکیب می‌شود. این توزیع احتمالی معرف عدم قطعیت در باره پارامتر پس از در نظر گرفتن داده‌ها است (Blasco, 2001). در رویکرد کلاسیک، پارامتر ثابت ولی ناشناخته در نظر گرفته می‌شود و تنها منبع تصادفی، توزیع داده‌ها است. تحت رویکرد بیزی، پارامترها تصادفی هستند و با کمک یک توزیع احتمال، توزیع پسین ارزیابی می‌شود. روش‌های MCMC برای نمونه‌برداری به طور مستقیم از توزیع پسین نمونه‌برداری نمی‌کنند. به جای آن یک زنجیر مارکف تنظیم می‌کنند که توزیع پسین را به‌صورت توزیع حد کننده دارا است. در الگوریتم Metropolis-Hasting، روش انجام این کار، نمونه‌بردار Gibbs و نمونه‌بردارهای دیگر هستند. به زنجیر مارکف به اندازه کافی اجازه داده می‌شود تا اجرا شود تا به توزیع حدکننده برسد. مقادیر به‌دست‌آمده از زنجیر مارکف در

جدول ۷- خلاصه‌های پسین پس از اعمال انقباض، احتمالات پسین در بر گرفتن در مدل و برآوردهای بی‌زی آثار مدل عملکرد دانه در تلاقی دوم

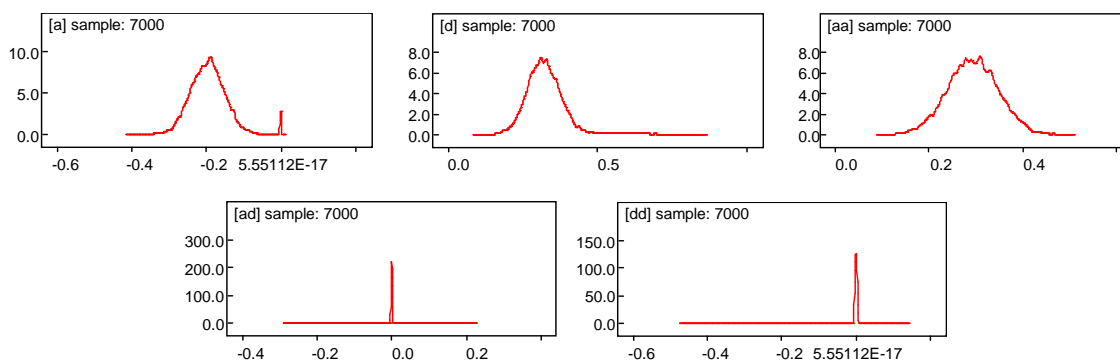
Table 7. Posterior summaries after applying shrinkage, posterior inclusion probabilities and Bayesian estimations of grain yield model effects in cross-2

Grain yield	Parameter	Shrinkage			Posterior inclusion probabilities	Bayesian estimation
		Mean	Standard deviation	MC error		
First year (Normal)	m	-	-	-	-	-0.653
	[a]	-0.647	0.105	0.007	1.0	-0.998
	[d]	1.458	0.169	0.010	1.0	14.440
	[aa]	0.755	0.081	0.004	1.0	2.584
	[ad]	-0.053	0.106	0.008	0.238	-
	[dd]	-0.930	0.136	0.008	1.0	-16.580
(First year (Stress))	m	-	-	-	-	-0.940
	[a]	-0.473	0.080	0.003	1.0	-0.672
	[d]	1.575	0.123	0.005	1.0	13.740
	[aa]	0.801	0.060	0.002	1.0	2.408
	[ad]	-0.291	0.081	0.003	0.989	-1.825
	[dd]	-0.880	0.099	0.004	1.0	-13.810
Second year (Normal)	m	-	-	-	-	1.256
	[a]	-0.544	0.055	0.001	1.0	-0.406
	[d]	0.001	0.010	2.48E-4	0.014	-
	[aa]	0.237	0.045	6.41E-4	1.0	0.371
	[ad]	-0.007	0.037	0.001	0.053	-
	[dd]	0.002	0.016	4.57E-4	0.30	-
Second year (Stress)	m	-	-	-	-	1.199
	[a]	-0.043	0.101	0.008	0.189	-
	[d]	-7.86E-5	0.003	3.99E-5	0.003	-
	[aa]	8.88E-4	0.008	1.86E-4	0.014	-
	[ad]	-0.481	0.100	0.007	0.991	-2.006
	[dd]	1.23E-4	0.004	5.81E-5	0.005	-



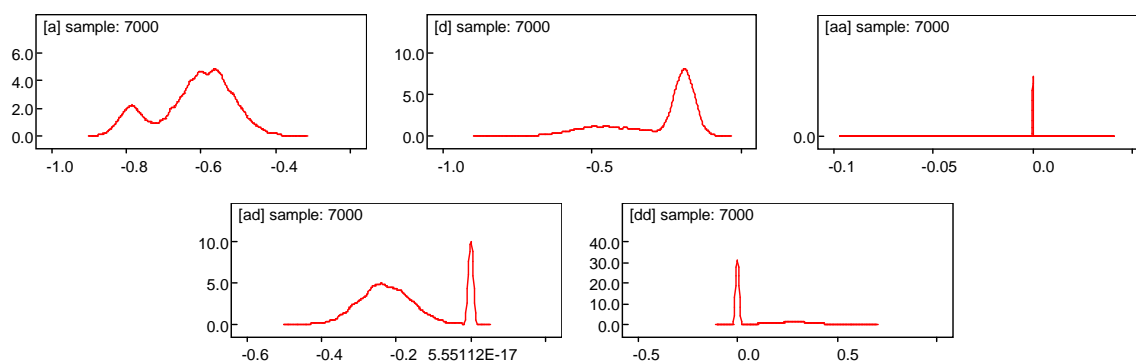
شکل ۱- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی اول در سال اول تحت شرایط آبیاری مطلوب

Figure 1. Posterior distribution densities of grain yield model effects after applying shrinkage and their posterior inclusion probabilities for cross-1 in 2015 under normal irrigation conditions



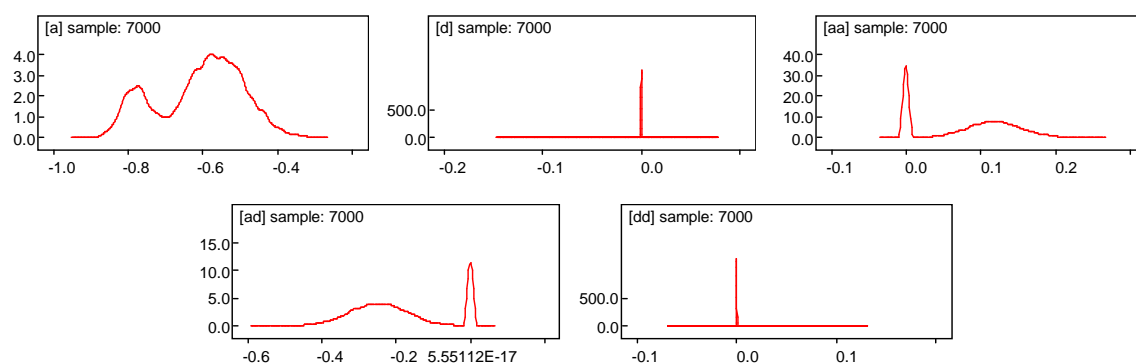
شکل ۲- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی اول در سال اول تحت شرایط تنش کم‌آبی

Figure 2. Posterior distribution densities of grain yield model effects after applying shrinkage and their posterior inclusion probabilities for Cross-1 in 2015 under water deficit conditions



شکل ۳- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی اول در سال دوم تحت شرایط آبیاری مطلوب

Figure 3. Posterior distribution densities of grain yield model effects after applying shrinkage and their posterior inclusion probabilities for Cross-1 in 2016 under normal irrigation conditions



شکل ۴- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی اول در سال دوم تحت شرایط تنش کم‌آبی

Figure 4. Posterior distribution densities of grain yield model effects after applying shrinkage and their posterior inclusion probabilities for Cross-1 in 2016 under water deficit conditions

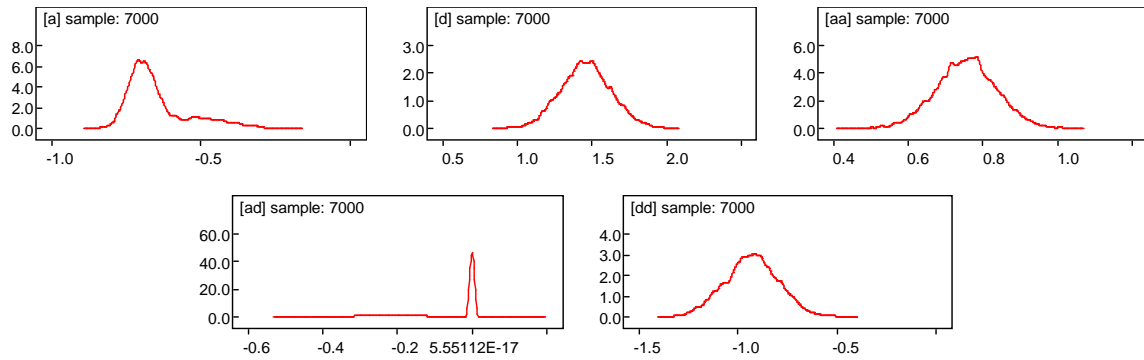
روش در اکثر موارد مشابه نبودند. روش LASSO بی‌زی نسبت به روش متداول اطلاعات بیشتری ارائه داد. توزیع پسین آثار غیرمهم مدل به سمت صفر منقبض شدند و از مدل‌ها کنار گذاشته شدند. در این روش احتمالات پسین در بر گرفتن آثار مدل نیز ارائه شد و انتخاب مهم‌ترین آثار در مدل‌های برازش یافته بر اساس یک معیار دقیق امکان‌پذیر شد. همچنین در سال اول تعداد پارامترهای مدل معادل تعداد نسل‌های ارزیابی شده بودند که استفاده از آزمون مقیاس مشترک را محدود می‌سازد، ولی این محدودیت در استفاده از تحلیل بی‌زی وجود نداشت و امکان کسب برآوردها برای تمام آثار وجود داشت. افزون بر این، روش بی‌زی خطاهای استاندارد کوچک‌تری داشت که به دلیل استفاده از اطلاعات بیشتر در باره پارامترهای مدل است که توسط توزیع پیشین فراهم می‌شود.

نتایج نشان‌دهنده اهمیت آثار اصلی و اپیستازی در توارث صفت عملکرد دانه است و باید آن‌ها را در زمان طراحی یک برنامه اصلاحی جدید برای بهبود جمعیت‌های گندم در نظر گرفت. برای توارث عملکرد دانه در گندم نقش آثار افزایشی (Novoselovic et al., 2004)، غیرافزایشی (Chowdhry et al., 1992; Munir et al., 2007)، آثار افزایشی و غالبیت (Fotokian et al., 2008) و آثار متقابل غیرآلی (Novoselovic et al., 2004; Khattab et al., 2013; Ijaz and Kashif, 2010) گزارش شده است. توارث عملکرد دانه در اکثر مدل‌ها تحت کنترل آثار ژنتیکی افزایشی و غیرافزایشی بود. هنگامی که آثار افزایشی بیشتر از غیرافزایشی هستند، گزینش در نسل‌های در حال تفرق ابتدایی موثر خواهد بود در حالی که اگر بخش غیرافزایشی بیشتر باشد، بهبود صفات نیاز به گزینش شدید در نسل‌های پیشرفته دارد. این پیشنهاد به این دلیل مناسب است که گزینش موثر در نسل‌های اولیه فقط زمانی تحقق می‌یابد که آثار ژنی افزایشی زیاد باشند و آثار محیطی کوچک باشند. در مواردی که آثار [ad] و [dd] معنی‌دار هستند، بهره‌برداری از این آثار از طریق گزینش در نسل‌های اولیه در حال تفرق امکان‌پذیر نیست (Kearsey and Pooni, 1998) و مانند اثر غالبیت در صورت تولید ارقام هیبرید می‌توان از آن‌ها استفاده کرد.

اگر چه روش حداقل مربعات برای برآورد پارامترها در تجزیه میانگین نسل‌ها کیفیت می‌کند، ولی برآوردهای بی‌زی در مقام مقایسه خطاهای استاندارد بسیار کوچک‌تری دارند که نشان‌دهنده قدرت بیشتر روش بی‌زی است. استفاده از اطلاعات بیشتر که توسط توزیع پیشین فراهم می‌شود، خطاهای استاندارد کوچکی را برای برآوردهای بی‌زی در پی دارد (Mettle et al., 2016). تیبشیرانی (Tibshirani, 1996) نشان داد که برآوردهای حداقل مربعات به دلیل دقت پیش‌بینی کم راضی‌کننده نیستند. برآوردهای حداقل مربعات متداول که توسط به حداقل رساندن مربع خطای باقی‌مانده به دست می‌آیند اغلب اریب کمی دارند ولی از واریانس بزرگی برخوردار هستند. وی نشان داد که دقت پیش‌بینی را می‌توان توسط اعمال انقباض به ضرایب مدل بهبود بخشید و با انجام این کار اریب کمی برای کاهش واریانس مقادیر پذیرفته می‌شود و از این رو دقت پیش‌بینی کل بهبود می‌یابد.

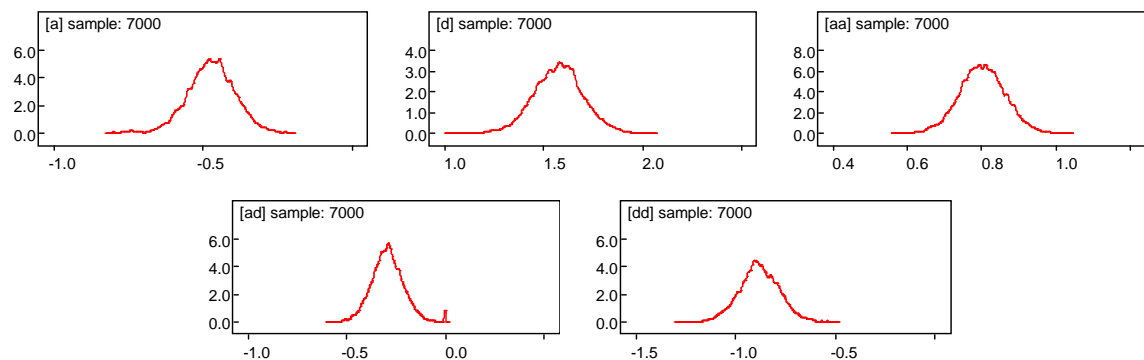
LASSO یک روش موثر و سریع برای اعمال انقباض و گزینش مدل است. مزیت اصلی تمام روش‌های انقباض این حقیقت است که آن‌ها را می‌توان به‌طور مستقیم در مدل کامل به کار برد و به جستجوی مدل نیازی نیست. ضرایب متغیرهای با آثار ضعیف بر متغیر پاسخ به سرعت به سمت صفر منقبض و بنابراین از ساختار مدل حذف می‌شوند. در حالی که انقباض کم‌تری برای ضرایب مهم مدل به کار برده می‌شود (Lykou and Ntzoufras, 2013). در تجزیه LASSO یکی از مسائل مهم، انتخاب مقدار پارامتر انقباض λ است. این پارامتر در واقع کل فرآیند را کنترل می‌کند. اگر یک مقدار کم برای λ انتخاب شود، هیچ انقباضی (و بنابراین هیچ گزینشی) انجام نمی‌شود. در حالی که اگر این مقدار بسیار زیاد باشد، تمام ضرایب به سمت صفر منقبض می‌شوند (Lykou and Ntzoufras, 2013). بنابراین انتخاب نادرست ممکن است موجب فوق انقباض ضرایب شود یا تناقض Lindley-Bartlett (Lindley, 1957; Bartlett, 1957) را فعال کند.

مقایسه برآوردهای آثار مدل در روش متداول و روش بی‌زی نشان داد که برآوردهای این دو روش نزدیک یکدیگر بودند. با وجود این، معرفی بهترین مدل‌ها توسط این دو



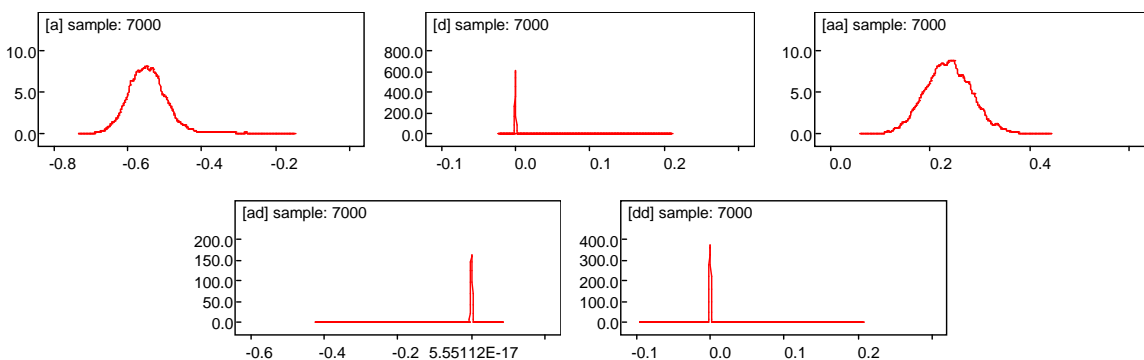
شکل ۵- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی دوم در سال اول (آبیاری مطلوب)

Figure 5. Posterior distribution densities of grain yield model effects after applying shrinkage and their posterior inclusion probabilities for Cross-2 in 2015 under normal irrigation conditions



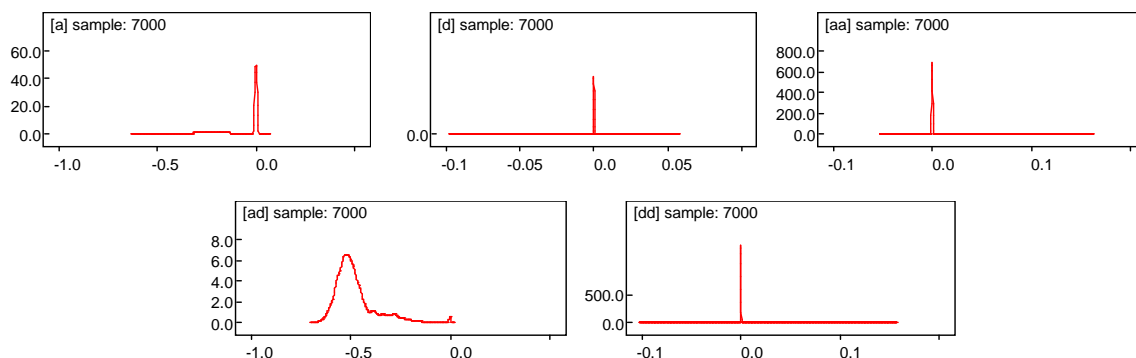
شکل ۶- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی دوم در سال اول تحت شرایط تنش

Figure 6. Posterior distribution densities of grain yield model effects after applying shrinkage and their posterior inclusion probabilities for Cross-2 in 2015 under water deficit conditions



شکل ۷- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی دوم در سال دوم تحت شرایط آبیاری مطلوب

Figure 7. Posterior distribution densities of grain yield model effects after applying shrinkage and their posterior inclusion probabilities for Cross-2 in 2016 under normal irrigation condition



شکل ۸- چگالی‌های توزیع پسین آثار موجود در مدل عملکرد دانه پس از اعمال انقباض و احتمالات پسین در بر گرفتن آن‌ها در مدل برای تلاقی دوم در سال دوم تحت شرایط تنش

Figure 8. Posterior distribution densities of grain yield model effects after applying shrinkage and posterior inclusion for model effects for Cross-2 in 2016 under water deficit conditions

نتیجه‌گیری کلی

مقایسه برآوردهای حاصل از روش متداول و روش بی‌زی در این تحقیق نشان داد که برآوردهای هر دو روش نزدیک به هم بودند، اما معرفی بهترین مدل‌ها توسط این دو روش در بیش‌تر موارد مشابه نبود. روش LASSO بی‌زی نسبت به روش متداول اطلاعات بیش‌تری ارائه کرد. در این روش احتمالات پسین در بر گرفتن آثار مدل نیز ارائه و انتخاب مهم‌ترین آثار در مدل‌های برازش‌یافته بر اساس یک معیار دقیق امکان‌پذیر شد. افزون بر این، روش بی‌زی خطاهای استاندارد کوچک‌تری داشت که به دلیل استفاده از اطلاعات بیش‌تر درباره پارامترهای مدل بود که توسط توزیع پیشین فراهم می‌شود. نتایج روش بی‌زی در مورد توارث عملکرد دانه گندم نیز اهمیت آثار اصلی و اپیستازی را نشان داد که باید آن‌ها را در بهبود جمعیت‌های گندم در نظر گرفت.

برآوردهای معنی‌دار برای آثار غالبیت و برهمکنش از نوع غالبیت برای اکثر مدل‌ها نشان‌دهنده نقش آثار ژنی غیرقابل تثبیت در کنترل صفت بود. در این شرایط شدت گزینش باید در نسل‌های اولیه ملایم و در نسل‌های پیشرفته شدید باشد تا بتواند پیشرفت از طریق گزینش را بهبود بخشد. از این‌رو روش‌هایی که آثار غیرافزایشی و برهمکنش غیرآلی را به‌کار می‌برند، مثل تولید هیبرید (در صورت رفع موانع تکنیکی اجرای آن)، می‌توانند برای بهبود ژنتیکی عملکرد امیدبخش باشند. همچنین، چند دوره گزینش دوره‌ای و به‌دنبال آن اجرای روش شجره‌ای نیز می‌تواند برای کاربرد موثر هم‌زمان هر سه نوع آثار ژنی مفید باشد. این امر منجر به افزایش تنوع در نسل‌های پیشرفته برای گزینش موثر با حفظ هتروزیگوسیتی قابل توجه از طریق تلاقی بوته‌های گزینش شده در نسل‌های در حال تفرق اولیه خواهد شد.

References

- Balestre, M., Von Pinho, R. G. and Brito, A. H. 2012. Bayesian inference to study genetic control of resistance to gray leaf spot in maize. *Genetics and Molecular Research* 11 (1): 17-29.
- Bartlett, M. S. 1957. Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)* 120 (1): 48-70.
- Blasco, A. 2001. The Bayesian controversy in animal breeding. *Journal of Animal Science* 79 (8): 2023-2046.
- Chowdhry, M. A., Rafiq, M. and Alam, K. 1992. Genetic architecture of grain yield and certain other traits in bread wheat. *Pakistan Journal of Agricultural Research* 13 (3): 216-220.
- Dellaportas, P., Forster, J. J. and Ntzoufras, I. 2002. On Bayesian model and variable selection using MCMC. *Statistics and Computing* 12 (1): 27-36.
- Fotokian, M. H., Ahmadi, J. and F. Orang, S. 2008. Genetic assay of some traits in wheat (*Triticum aestivum* L.) under drought stress conditions using generation mean analysis. *Iranian Journal of Biology* 22 (3): 431-441. (In Persian with English Abstract).
- Gelfand, A. E. and Smith, A. F. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85 (410): 398-409.

- Geman, S. and Geman, D. 1984.** Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721-741.
- Gomez, K. A. and Gomez, A. A. 1984.** Statistical procedures for agricultural research. John Wiley & Sons.
- Hastings, W. K. 1970.** Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57 (1): 97-109.
- Ijaz, U. S. and Kashif, M. 2013.** Genetic study of quantitative traits in spring wheat through generation means analysis. *American-Eurasian Journal of Agricultural and Environmental Sciences* 13 (2): 191-197.
- Kearsey, M. J. and Pooni, H. S. 1998.** The genetical analysis of quantitative traits. Stanley Thornes Publishers, Ltd.
- Khattab, S. A. M., Esmail, R. M. and Al-Ansary, A. M. F. 2010.** Genetical analysis of some quantitative traits in bread wheat (*Triticum aestivum* L.). *New York Science Journal* 3: 152-157.
- Kuo, L. and Mallick, B. 1998.** Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*: 65-81.
- Lindley, D. V. 1957.** A statistical paradox. *Biometrika* 44 (1-2): 187-192.
- Lykou, A. and Ntzoufras, I. 2011.** WinBUGS: A tutorial. *Wiley Interdisciplinary Reviews: Computational Statistics* 3 (5): 385-396.
- Lykou, A. and Ntzoufras, I. 2013.** On Bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing* 23: 361-390.
- Lynch, S. M. 2007.** Introduction to applied Bayesian statistics and estimation for social scientists. Springer Science and Business Media.
- Mather, K. and Jinks, J. L. 1971.** Biometrical genetics. Cornell University Press, Ithaca, N.Y.
- Mathew, B., Bauer, A. M., Koistinen, P., Reetz, T. C., Léon, J. and Sillanpää, M. J. 2012.** Bayesian adaptive Markov Chain Monte Carlo estimation of genetic parameters. *Heredity* 109 (4): 235-245.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. 1953.** Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21: 1087-1092.
- Mettle, F. O., Asiedu, L., Quaye, E. N. and Asare-Kumi, A. A. 2016.** Comparison of least squares method and Bayesian with multivariate normal prior in estimating multiple regression parameters. *British Journal of Mathematics and Computer Science* 15 (1): 1-8.
- Munir, M., Chowdhry, M. A. and Ahsan, M. 2007.** Generation means studies in bread wheat under drought condition. *International Journal of Agriculture and Biology* 9 (2): 282-286.
- Nezhadahmadi, A., Prodhon, Z. H. and Faruq, G. 2013.** Drought tolerance in wheat. *The Scientific World Journal* 2013: 1-12.
- Novoselovic, D., Baric, M., Drezner, G., Gunjaca, J. and Lalic, A. 2004.** Quantitative inheritance of some wheat plant traits. *Genetics and Molecular Biology* 27 (1): 92-98.
- Ntzoufras, I. 2011.** Bayesian modeling using WinBUGS. Vol. 698. John Wiley & Sons.
- SAS Institute. 2002.** SAS user's guide: Statistics version 9 for windows. SAS Institute., Carry, NC.
- Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D. 2003.** WinBUGS user manual. MRC Biostatistics Unit, Cambridge.
- Tibshirani, R. 1996.** Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B (Methodological)*: 267-288.
- Waldmann, P., Hallander, J., Hoti, F. and Sillanpää, M. J. 2008.** Efficient Markov Chain Monte Carlo implementation of Bayesian analysis of additive and dominance genetic variances in noninbred pedigrees. *Genetics* 179 (2): 1101-1112.
- Wang, D., El-Basyoni, I. S., Baenziger, P. S., Crossa, J., Eskridge, K. M. and Dweikat, I. 2012.** Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. *Heredity* 109 (5): 313-319.
- Wang, D., Eskridge, K. M. and Crossa, J. 2011.** Identifying QTLs and epistasis in structured plant populations using adaptive mixed LASSO. *Journal of Agricultural, Biological and Environmental Statistics* 16 (2): 170-184.
- Xu, S. 2003.** Estimating polygenic effects using markers of the entire genome. *Genetics*, 163(2): 789-801.
- Xu, S. 2007.** An empirical Bayes method for estimating epistatic effects of quantitative trait loci. *Biometrics* 63 (2): 513-521.



Bayesian inference to study genetic control of water deficit stress tolerance in wheat by LASSO method

Seyyede Fatemeh Danyali^{1*}, Parviz Safari² and Mehdi Rahimi³

Received: June 10, 2017

Accepted: February 26, 2018

Abstract

Drought is the main abiotic stress seriously influencing wheat production and quality in Iran. Information about genetic controlling drought tolerance inheritance is necessary to determine the type of breeding program as well as develop tolerant cultivars. In this study, Bayesian inference using LASSO method used to identify the most important gene effects related to drought tolerance in context generation mean analysis. For this purpose, field experiments consist of two pairs of crosses with non-tolerant and tolerant cultivars and generations derived from them were carried out across two years as split plot designs based on RCBD with three replications in which main plots assigned to irrigation treatment consist of two levels (well watered and cessation of irrigation at pollination stage) and sub-plots given to the generations. Bayesian inference is an alternative approach which combines available prior knowledge (prior distribution) with the information contained in the data. The result is the posterior distribution containing all information to interpret genetic structure. LASSO is an effective method to apply shrinkage and selection on model variables. Non-important effects in the model shrunk toward zero and excluded from the model. While for important effects, less shrinkage will be achieved. Since the additive, dominance and epistatic gene actions involved in drought tolerance inheritance, methods which utilize all type of gene effects, like recurrent selection followed by pedigree method may be useful for drought tolerance stress improvement.

Keywords: Additive effect, Dominance effect, Epistatic effect, Gene action, Generation mean analysis

1. Ph. D. Candidate, Dept. of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

2. Graduated Ph. D., Dept. of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

3. Assist. Prof., Dept. of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, Iran

* Corresponding author: danyalisf@gmail.com