

## اثر توزیع‌های پیشین مدل‌های آماری متفاوت بر صحت پیش‌بینی ژنومی: مطالعه شبیه‌سازی

- یحیی محمدی (نویسنده مسئول)  
استادیار و عضو هیات علمی دانشگاه ایلام.
- جواد احمدپناه  
عضو هیات علمی، بخش علوم دامی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرمانشاه، ایران.
- حسن بانه  
استادیار، موسسه تحقیقات علوم دامی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، کرج، ایران.

تاریخ دریافت: دی ۱۳۹۸ تاریخ پذیرش: تیر ۱۳۹۹

شماره تماس نویسنده مسئول: ۰۹۱۸۳۴۲۱۴۶۴

Email: mohamadi\_yahya@yahoo.com

شناسه دیجیتال (DOI): 10.22092/asj.2020.341337.2024

### چکیده

در انتخاب ژنومی از نشانگرهای SNP در کل سطح ژنوم برای برآورد اثرات نشانگر استفاده می‌شود. به کمک روش‌های مختلف آماری، ارزش‌های اصلاحی ژنومی برای حیوانات تخمین زده می‌شود. در پژوهش کنونی صحت ارزش‌های اصلاحی ژنومی مستقیم حیوان به کمک روش‌های Bayes-A، B-LASSO gamma، B-LASSO beta، BGLR و برای برآورد اثرات نشانگر SNP در دو مقدار وراثت‌پذیری ۰/۳ و ۰/۵ مقایسه شدند. میزان پارامترهای پیش‌تنظیمی  $\pi$  (نسبت نشانگرهای SNP که با واریانت‌های علی در عدم تعادل پیوستگی (LD) قرار دارند) برای سه راهبرد ۰/۱، ۰/۳ و ۰/۵ و  $s^2$  (پارامتر مقیاس میانگین پیشین) برای چهار راهبرد ۰/۱، ۰/۳، ۱۰ و ۱۰۰ شبیه‌سازی و تاثیر آن‌ها بر صحت پیش‌بینی ژنومی بررسی گردید. بیشترین صحت‌های ارزش‌های اصلاحی ژنومی مستقیم با روش Bayes-A برای وراثت‌پذیری ۰/۳ به مقدار ۰/۸۸ و برای وراثت‌پذیری ۰/۵ به مقدار ۰/۶۹ برآورد شدند. ضریب رگرسیون برای تمام روش‌های برآورد اثرات نشانگر در وراثت‌پذیری ۰/۳ نسبت به وراثت‌پذیری ۰/۵ اریبی کمتر را نشان داد. در وراثت‌پذیری ۰/۳ کمترین مقدار خطای آزمایشی مربوط به روش Bayes-A (۱۲۱/۲) و بیشترین مقدار آن مربوط به روش B-LASSO (۱۶۵/۲) بود. بیشترین مقدار صحت مربوط به Bayes-A در  $\pi$  برابر ۰/۵ (۰/۸۱) و کمترین مقدار صحت در  $\pi$  برابر با ۰/۱ (۰/۴۵) در روش BGLR دیده شد. میزان صحت پیش‌بینی ژنومی با تغییر پارامتر  $s^2$  از ۰/۱ به سمت ۱۰ و ۱۰۰ کاهش بیشتری داشت. نتایج پژوهش کنونی نشان داد که میزان بهینه پارامتر پیش‌تنظیمی  $\pi$  برای حداکثر نمودن صحت پیش‌بینی ژنومی برای صفات با وراثت‌پذیری متوسط به بالا (۰/۳) از ۰/۴۱ تا ۰/۵۶ و برای صفاتی با وراثت‌پذیری پایین (۰/۰۵) از ۰/۵۶ تا ۰/۷۳ پیشنهاد می‌گردد. همچنین مقدار بهینه پارامتر پیش‌تنظیمی  $s^2$  برای صفات با وراثت‌پذیری متوسط به بالا (۰/۳) از ۰/۰۵ تا ۰/۱۲ و برای صفات با وراثت‌پذیری پایین (۰/۰۵) از ۰/۰۲ تا ۰/۰۵ پیشنهاد می‌گردد.

واژه‌های کلیدی: پارامترهای پیش‌تنظیمی، توزیع اثرات نشانگر، صحت پیش‌بینی، شبیه‌سازی.

Animal Science Journal (Pajouhesh & Sazandegi) No 130 pp: 191-202

### The effect of prior distributions of different statistical models on the accuracy of genomic prediction: simulation study.

By: Yahya Mohammadi<sup>\*1</sup>, Javad Ahmadpanah<sup>2</sup>, Hassan Baneh<sup>3</sup>

1: Assistant professor, Department of Animal Science, Faculty of Agriculture, Ilam University, Ilam, Iran

2: Animal Science Research Department, Kermanshah Agriculture and Natural Resources Research and Education Center, Kermanshah, Iran.

3: Animal Science Research Institute of Iran, Agriculture Research, Education and Extension Organization (AREEO), Karaj, Iran.

Received: February 2020

Accepted: July 2020

In the genomic selection, SNP markers across whole genome are used to estimate the marker effects. Genomic breeding value of animals can be predicted by different statistical methods. Genomic breeding value of animals can be predicted by different statistical methods. In the present study, accuracies of the predicted direct genomic breeding values were compared under several statistical models including Bayes A, B-LASSO gamma, B-LASSO beta and BGLR by considering two heritabilities of 0.3 and 0.05. Three values of  $\pi$  (0.1, 0.3, and 0.5) and four values of  $s^2$  (0.01, 0.1, 10, and 100) were simulated and correspondingly evaluated in terms of accuracy. The obtained results showed the highest accuracy of direct genomic breeding value was obtained under Bayes-A method, which were 0.88 and 0.69 for heritabilities of 0.3 and 0.05, respectively. Regression coefficients of methods for estimating the marker effects were more unbiased under heritability of 0.3 than 0.05. By considering the heritability of 0.3, the lowest and highest error were obtained under Bayes-A (121.2) and B-LASSO beta (165.2) methods, respectively. Under BGLR method, the highest and lowest accuracy of Bayes-A were obtained for  $\pi$ , 0.5 (0.81) and  $\pi$ , 0.1 (0.45). By increasing  $s^2$  parameter a decrease in the accuracy of genomic predictions was obtained. The obtained results suggested that to maximize the accuracy of the genetic prediction for traits with moderate to high heritability (0.3) optimal point of parameter  $\pi$  may ranged from 0.41 to 0.56 and while for traits with low heritability from 0.56 to 0.73. Also, the optimal level of  $s^2$  hyperparameters parameter, for traits with medium to high heritability (0.3) ranged from 0.05 to 0.12 and for traits with low heritability (0.05) from 0.02 to 0.5.

**Key words:** Hyperparameters, Distribution of marker effects, Predicting genomic accuracy, Simulation.

#### مقدمه

(QTL) می‌باشد (Goddard و Hayes، ۲۰۰۷). برای محاسبه میانگین شرطی نیاز به یک توزیع پیشین<sup>۳</sup> برای اثرات QTL است. اما در عمل هم ژنوتیپ QTL ها و توزیع اثرات QTL ناشناخته می‌باشد، که از نشانگرهای SNP برای ردیابی QTL ها به کمک عدم تعادل پیوستگی استفاده می‌شود. در مطالعه‌ای از روش‌های بیزین با توزیع نرمال (BLUP) و توزیع t استوندنت (Bayes-A) و (Bayes-B) برای محاسبه اثرات نشانگرها به کمک داده شبیه-سازی استفاده گردید (Meuwissen و همکاران، ۲۰۰۱). نتایج

در سالیان اخیر استفاده از فن آوری ژنومی با تراشه‌های نشانگری با تراکم بالا باعث بهبود کارایی انتخاب در تعداد زیادی از دام‌های اهلی شده است (Scheffers و Weigel، ۲۰۲۰؛ Harris و همکاران، ۲۰۰۸). بهترین روش آماری برای تخمین ارزش ژنومی مستقیم حیوان<sup>۱</sup> تحت بررسی و مطالعه می‌باشد (Nicolazzi و همکاران، ۲۰۱۳). در مطالعه‌ای گزارش شده که بهترین روش برآورد ارزش ژنومی مستقیم حیوان محاسبه میانگین شرطی ارزش‌های اصلاحی ژنومی افراد در هر جایگاه صفات کمی<sup>۲</sup>

<sup>3</sup>- Prior distribution

<sup>1</sup>- Direct genomic values (DGV)  
<sup>2</sup>- Quantitative trait loci

همکاران، ۲۰۰۹). در کشورهای در حال توسعه با توجه به دسترسی پژوهشگران به تعداد کم حیوانات تعیین ژنوتیپ شده، دستیابی به حداکثر صحت پیش بینی ژنومی برای صفات با وراثت-پذیری کم و زیاد نیاز به روش های آماری مناسب برای برآورد اثرات نشانگرها می باشد. از طرف دیگر در راهبردهای مختلف شبیه سازی تنظیم دقیق پارامترهای پیش تنظیم بسیار مهم است، مشخصات بهینه پارامترهای پیش تنظیمی به چندین عامل اصلی از جمله پارامتر  $\pi$  (نسبت نشانگرهای SNP که با واریانت های علی<sup>۱۰</sup> در عدم تعادل پیوستگی (LD) قرار دارند) و پارامتر  $\gamma$  (درجه آزادی و  $S^2$  پارامتر مقیاس میانگین پیشین) بستگی دارند.

اهداف پژوهش کنونی، مقایسه صحت ارزش های اصلاحی ژنومی مستقیم حیوان به کمک روش های آماری متفاوت با سه توزیع متفاوت برای اثرات نشانگر SNP و همچنین مقایسه صحت برآورد ارزش های ژنومی مستقیم حیوان با توزیع متفاوت پارامترهای  $\pi$ ، درجه آزادی و پارامتر مقیاس  $S^2$  و تعیین مقدار بهینه این پارامترهای پیش تنظیمی بود.

### مواد و روش ها

از نرم افزار QMSim برای ایجاد یک جمعیت در طول زمان استفاده گردید (Sargolzaei و Schenkel، ۲۰۰۹). در ابتدا ۱۰۰۰ نسل مجزا با ۱۰۰۰۰ فرد شبیه سازی گردید. برای ایجاد LD، از نسل ۱۰۰۰ به نسل ۱۰۲۰ تعداد جمعیت به ۵۰۰۰ حیوان کاهش یافت. در ادامه به کمک افراد جمعیت پایه، ۸ نسل تلاقی صورت گرفت. افراد ۵ نسل اول بعنوان افراد جمعیت مرجع<sup>۱۱</sup> و افراد ۳ نسل آخر به عنوان افراد جمعیت تایید<sup>۱۲</sup> در نظر گرفته شدند. تعداد ۵۰ QTL که جایگاه آنها روی کروموزوم مشخص بود، با اثرات افزایشی طراحی گردید. برای توزیع اثرات QTL توزیع گاما در نظر گرفته شد. تعداد ۵۰۰۰ نشانگر SNP بطور مساوی بر روی ۴ کروموزوم شبیه سازی شدند. مدل فنوتیپی پیش از تصحیح برای افراد با اثرات ثابت به صورت ذیل بود:

مطالعه آنها نشان داد که Bayes-A و Bayes-B بهتر از BLUP اثرات نشانگر را برآورد می کنند. مطالعات دیگری با استفاده از داده های واقعی نشان داد که صحت ارزش های ژنومی مستقیم به کمک مدل های بیزین غیرخطی<sup>۴</sup> و BLUP بستگی به نوع صفت مورد بررسی دارد (Pintus و همکاران، ۲۰۱۳؛ Gaspa و همکاران، ۲۰۱۳). نتایج حاصل از مطالعه بر روی ۲۷ صفت در گاوهای شیری به کمک مدل های غیر خطی (مانند Bayes-A) و BLUP نشان داد که برای بعضی صفات نظیر درصد چربی و پروتئین شیر (به دلیل داشتن ژن های بزرگ اثر) صحت برآورد ارزش اصلاحی ژنومی به کمک مدل های غیرخطی بیشتر از BLUP است (VanRaden و همکاران، ۲۰۰۹). پژوهشگران با بررسی تفاوت بین صحت ارزش اصلاحی ژنومی داده های گاو شیری به کمک روش های BLUP و Bayes-A گزارش کردند که دامنه صحت از ۲ درصد برای متغیر طبقه بندی سود استرالیایی<sup>۵</sup> تا ۷ درصد برای درصد پروتئین شیر بود (Hayes و همکاران، ۲۰۰۹). از طرف دیگر در روش بیز LASSO یک رویکرد سلسه مراتبی<sup>۶</sup> برای محاسبه اثرات نشانگرها ترسیم شد بطوری که اثرات نشانگرهای SNP دارای یک توزیع نرمال با واریانس پیشین<sup>۷</sup> مشخص بود (Casella و Park، ۲۰۰۹). این واریانس دارای یک توزیع نمایی<sup>۸</sup> و پارامتر از پیش تنظیم شده<sup>۹</sup>  $\lambda$  است که از طریق داده ها ترسیم می شود. در روش های BLUP، Bayes-A و Bayes-B واریانس پیشین برای نشانگرهای SNP، ثابت و در بیز LASSO، ناشناخته است. از طرفی قابلیت اطمینان مدل های پیش بینی ژنومی براساس استفاده از نشانگرهای SNP به مشخصات مناسب پارامترهای پیش تنظیم شده بستگی دارد (Yang و همکاران، ۲۰۱۵). در یک مطالعه که با داده شبیه سازی انجام گردید، پنج نوع پارامتر و دو نوع توزیع (گاما و بتا) برای اطلاعات پیشین  $\lambda$  در نظر گرفته شد نتایج تحقیق آنها نشان داد که از لحاظ صحت برآورد اثرات نشانگرهای SNP تفاوت کمی بین روش ها وجود دارد (de los Compos و

<sup>1</sup>- Non-linear Bayesian model

<sup>2</sup>- Australian profit ranking

<sup>3</sup>- Hierarchical approach

<sup>4</sup>- Prior variance

<sup>5</sup>- Exponential distribution

<sup>6</sup>- Hyperparameter

<sup>7</sup>- Causal variants

<sup>1</sup>- Training population

<sup>2</sup>- Testing population

(۱) مدل

ارزش ژنومی واقعی حیوان به کمک معادله زیر برآورد گردید:

$$DGV_i = \mu + \sum_{i=1}^Z X_{ij} g_j \quad (۵) \text{ معادله}$$

در روش BGLR توزیع اثرات نشانگرها بصورت نرمال و واریانس این توزیع در هر تکرار به کمک نمونه‌گیری گیس محاسبه شد (Verbyla و همکاران، ۲۰۱۰). در روش Bayes-A توزیع اثرات نشانگر  $t$  استودنت و درجه آزادی برای معادلات ۴/۰۱ فرض شد. در روش‌های بیز LASSO (LASSO beta و gamma) برای توزیع اثرات نشانگر از فرمول زیر استفاده شد (de los Campos و همکاران، ۲۰۰۹).

معادله (۶)

$$\begin{aligned} \text{likelihood: } & p(y|\mu, g, u_i, \sigma^2) \\ & \prod N(y_i|\mu + xig + u_i + \sigma^2) \\ \text{prior: } & p(\mu, \mu, g, u_i, \sigma^2, t^2, \lambda^2) \\ & p(\mu|\sigma^2) p(g|\sigma^2, t^2) p(\sigma^2) p(u|\sigma^2) \end{aligned}$$

$$= N(\mu|0, \sigma^2) \{\prod_{j=1}^t N(g_j|0, \sigma^2, t_j)$$

$$\{\prod_{j=1}^t EXP(t_j^2(\lambda))\} G(\lambda^2|\alpha_1, \alpha_2)$$

در معادله ۶،  $N(y_i|\mu + xig + u_i + \sigma^2)$  یک توزیع نرمال با میانگین  $\mu + xig + u_i$  و واریانس  $\sigma^2$  است.  $N(\mu|0, A, \sigma^2)$ ،  $N(g_j|0, \sigma^2, t^2)$  و  $N(u|0, A, \sigma^2)$  به ترتیب میانگین کل، اثرات نشانگر SNP و ارزش اصلاحی ژنومی با میانگین صفر و واریانس‌های  $\sigma^2$ ،  $\sigma^2 t^2$  و  $A\sigma^2$  بود.  $X^{-2}(\sigma^2|dof_w, s_u)$  و  $X^{-2}(\sigma^2|dof_e, s_e)$  به ترتیب توزیع کای اسکور معکوس<sup>۱۴</sup> با درجه آزادی dof و پارامتر S برای واریانس باقیمانده و ژنومی است. همچنین  $EXP(t^2|\lambda)$  یک توزیع نمایی برای واریانس نشانگر که بوسیله پارامتر  $\lambda$  (پارامتر تنظیمی) کنترل می‌شود. در نهایت  $G(\lambda^2|\alpha_1, \alpha_2)$  یک توزیع گاما با پارامتر شکل و میزان به ترتیب برای  $\alpha_1, \alpha_2$  بود.

$$Y_{ijk} = \mu + SEX_i + GEN_{ij} + e_{ijk}$$

در معادله ۱،  $Y_{ijk}$  ارزش فنوتیپی فرد  $i$  در جمعیت مرجع،  $\mu$  میانگین کل افراد،  $SEX_i$  اثرات ثابت جنس (۱ و ۲)،  $GEN_{ij}$  اثرات ثابت نسل (از ۱ تا ۸) و  $e_{ijk}$  اثرات تصادفی برای حیوان می‌باشد. بعد از تصحیح فنوتیپ مدل بصورت زیر تغییر کرد:

مدل (۲)

برای تمام حیوانات ارزش اصلاحی حقیقی<sup>۱۳</sup> محاسبه گردید. صحت برآورد از همبستگی بین ارزش ارثی مستقیم حیوان (DGV) و ارزش اصلاحی حقیقی (TBV) محاسبه گردید. هر چه عدد ضریب رگرسیون پیش‌بینی ژنومی (که از ضریب رگرسیون ارزش اصلاحی واقعی بر ارزش اصلاحی مستقیم حیوان به دست می‌آید) نزدیک به یک باشد، نشان‌دهنده این است که مقیاس ارزش اصلاحی ژنومی واقعی با ارزش اصلاحی تخمینی یکسان و هم‌اندازه بوده و اریب آن نیز کمتر است. همچنین میانگین مربعات خطای پیش‌بینی ژنومی (MSE) که از رابطه زیر پیش‌بینی گردید (Howard و همکاران، ۲۰۱۴):

(۳)

معادله

$$MSE = \frac{1}{N} \sum (DGV - TBV)^2$$

### توصیف مدل‌ها برای اثرات نشانگر

مدل زیر بر روی داده‌ها برازش گردید:

$$y = 1_n \mu + Z_u u + X_g g + e \quad (۴) \text{ معادله}$$

در معادله ۴،  $y$  بردار فنوتیپی افراد در جمعیت مرجع،  $\mu$  میانگین جامعه،  $1_n$  بردار یک،  $X$  ماتریس طرح با عناصر ۰، ۱ و ۲ بترتیب برای ژنوتیپ‌های AA، AB و BB در حیوان  $i$  و نشانگر SNP  $j$ ،  $u$  بردار ارزش‌های اصلاحی ژنومی (توزیع نرمال برای روش BGLR،  $t$  استودنت برای Bayes-A و نمایی برای بیز LASSO)،  $Z$  ماتریس طرح که داده‌ها را به اثرات اصلاحی ژنومی ارتباط می‌دهد و  $e$  بردار تصادفی باقیمانده بود. همچنین

<sup>1</sup>- Inverted Chi- square

<sup>3</sup>- True breeding value

شناسایی ژنگاه‌ها موثر بر صفات را به دلیل عدم تعادل پیوستگی (LD) افزایش داده و متعاقب آن سهم بیشتری از واریانس ژنتیکی برای این صفات قابل توجه، و در نتیجه صحت انتخاب ژنومی افزایش می‌یابد (de los Campos و همکاران، ۲۰۱۳).

در معادله ۶ برای توزیع‌های شرطی از نمونه‌گیری گیبس<sup>۱۵</sup> برای پاسخ به معادلات استفاده شد. پارامتر شکل و مقیاس برای پارامتر  $\lambda$  تنظیمی بترتیب به صورت  $G(\lambda^2|P=0.1, S=1 \times 10^{-0.4})$  تعریف شد. از بیز LASSO بنا بدلیل انعطاف‌پذیری بیشتر پارامتر تنظیمی  $\lambda$  برای برآورد اثرات نشانگر استفاده شد. اطلاعات پیشین برای اثرات نشانگر در بیز beta بصورت مسطح<sup>۱۶</sup> با دامنه وسیعتر نسبت به بیز LASSO گاما است. در این توزیع پارامتر  $\lambda$  تنظیمی به صورت  $\alpha \text{Beta}(\lambda|max=400, \alpha_1=1.4, \alpha_2=1.4)$  تعریف شد. در مجموع از یک زنجیره با طول ۱۰۰۰۰ تکرار استفاده گردید. همچنین ۱۵۰۰ نمونه اول به عنوان دوره گرم کردن<sup>۱۷</sup> در نظر گرفته شد. دو مقدار برای وراثت‌پذیری شامل ۰/۳ (متوسط به بالا) و ۰/۰۵ (پایین) در نظر گرفته شد. در مرحله بعد مقدار پارامتر  $\pi$  برای سه راهبرد ۰/۳، ۰/۵ و ۰/۱، مقدار پارامتر  $S^2$  در دامنه وسیعی از ۰/۱، ۱۰ و ۱۰۰ و پارامتر  $\gamma$  با درجه آزادی ۱۵۰ برای تاثیر این سه پارامتر بر صحت پیش‌بینی ژنومی شبیه‌سازی شد. سپس برای بهترین مقدار بهینه پارامترهای پیش‌تنظیمی ( $S^2$  و  $\pi$ ) آنالیز آماری معادلات رگرسیونی صورت گردید.

## نتایج و بحث

در جدول ۱ صحت، ضریب رگرسیون و میانگین مربعات خطای آزمایشی نشان داده می‌شود. مطابق جدول ۱ در تمام روش‌های برآورد اثرات نشانگر با افزایش وراثت‌پذیری صحت پیش‌بینی ژنومی افزایش یافته و مقدار این افزایش نیز معنی‌دار بدست آمد ( $P \leq 0.01$ ). در یک مطالعه تاثیر وراثت‌پذیری صحت بر صحت پیش‌بینی ژنومی در داده‌های گاو و خوک نشان داد که با افزایش وراثت‌پذیری صحت پیش‌بینی ژنومی افزایش پیدا می‌کند که با نتایج پژوهش کنونی مطابقت دارد (Zhang و همکاران، ۲۰۱۹). در صفاتی با وراثت‌پذیری پایین احتمالاً به دلیل کاهش تعداد ژنگاه‌های تأثیرگذار، می‌توان تعداد نشانگرها در سطح ژنوم را افزایش داد. افزایش تعداد نشانگرها، احتمالاً

<sup>2</sup>- Gibbs sampler

<sup>3</sup>- Flat prior

<sup>4</sup>- Burn-in

جدول ۱- صحت، ضریب رگرسیون و میانگین مربعات خطای آزمایشی برای روش‌های متفاوت برآورد اثرات نشانگر با داده شبیه‌سازی

BGLR	B-LASSO beta	B-LASSO gamma	Bayes-A	وراثت پذیری	
۰/۷۹ <sup>b</sup>	۰/۸۵ <sup>ab</sup>	۰/۸۶ <sup>ab</sup>	۰/۸۸ <sup>a</sup>	۰/۳	صحت*
۰/۵۹ <sup>b</sup>	۰/۶۷ <sup>ab</sup>	۰/۶۴ <sup>ab</sup>	۰/۶۹ <sup>a</sup>	۰/۰۵	
۱/۲۵ <sup>a</sup>	۰/۹۲۵ <sup>b</sup>	۰/۹۳۵ <sup>b</sup>	۱/۱۴ <sup>a</sup>	۰/۳	ضریب تابعیت**
۱/۴۴ <sup>a</sup>	۰/۸۷۷ <sup>b</sup>	۰/۸۹۷ <sup>b</sup>	۱/۳۲ <sup>a</sup>	۰/۰۵	
۱۴۳/۲ <sup>b</sup>	۱۶۵/۲ <sup>a</sup>	۱۶۷/۲ <sup>a</sup>	۱۲۱/۲ <sup>c</sup>	۰/۳	MSE
۱۴۹/۸ <sup>b</sup>	۱۶۹/۷ <sup>a</sup>	۱۷۳/۸ <sup>a</sup>	۱۳۱/۸ <sup>c</sup>	۰/۰۵	

- حروف متفاوت در هر ردیف صحت انتخاب، نشان دهنده معنی داری بین روش‌ها به کمک آزمون دانکن را نشان می‌دهد (P<0.05).

$$b = cov(TBV, DGV)/var(DGV) \quad ** \quad r = corr(TBV, DGV) *$$

اثر طراحی گردیدند، لذا انتظار افزایش صحت پیش‌بینی اثرات نشانگر توسط این روش نسبت به روش‌های B-LASSO قابل پیش‌بینی بود. نتایج پژوهش کنونی نیز در تطابق با این فرض قرار داشت. در مطالعه‌ای شبیه‌سازی به دلیل در نظر گرفتن QTL ها بزرگ اثر تفاوت صحت پیش‌بینی اثرات نشانگر در روش - Bayes-A نسبت به روش‌های B-LASSO معنی‌دار و زیاد گزارش شد (Xu و Yi، ۲۰۰۸). Nicolazzi و همکاران (۲۰۱۳) در پژوهشی صحت پیش‌بینی ژنومی برای صفات درصد چربی و درصد پروتئین شیر در گاوهای هلشتاین - فریزین استرالیایی را با کمک چهار روش برآورد اثر نشانگر (Bayes A, B-LASSO gamma, B-LASSO beta و BGLR) بررسی کردند. صحت پیش‌بینی ژنومی برای صفت درصد چربی شیر که تحت تاثیر ژن بزرگ اثر DGAT1 است در این روش‌ها به ترتیب ۰/۷۵، ۰/۷۱، ۰/۷۱ و ۰/۵۹ و برای صفت پروتئین شیر به ترتیب ۰/۵۲، ۰/۴۸، ۰/۴۸ و ۰/۴۷ گزارش شد. برای پاسخ به این سوال که چرا Bayes-A در مقایسه با B-LASSO فقط با دارا بودن یک نشانگر بزرگ اثر صحت بیشتری را برای اثرات نشانگر برآورد می‌کند، می‌توان گفت که توزیع پیشین در روش Bayes-A از انعطاف بیشتری برخوردار بوده و همچنین تمام اثرات جهش بوسیله LD بالای بین نشانگرهای SNP به دست می‌آید، در حالیکه در روش‌های B-LASSO توزیع پیشین بین نشانگرهای زیادی با اثرات کوچک توزیع می‌گردد (ter Braak و

مطابق جدول ۱ بیشترین صحت‌های ارزش اصلاحی ژنومی مستقیم حیوانات در روش Bayes-A و وراثت‌پذیری ۰/۳ به مقدار ۰/۸۸ و در وراثت‌پذیری ۰/۰۵ به مقدار ۰/۶۹ برآورد شدند. صحت ارزش اصلاحی ژنومی مستقیم حیوانات در روش B-LASSO gamma و وراثت‌پذیری ۰/۳ و ۰/۰۵ به ترتیب ۰/۸۶ و ۰/۶۴، در روش B-LASSO beta به ترتیب ۰/۸۵ و ۰/۶۷ و در روش BGLR به ترتیب ۰/۷۹ و ۰/۵۹ بدست آمدند. صحت پیش‌بینی ژنومی روش Bayes-A نسبت به دیگر روش‌ها تا حدودی بیشتر بود. متوسط صحت بدست آمده برای روش Bayes-A نسبت به روش BGLR، ۹ واحد افزایش بیشتر و معنی‌داری را نشان داد (P≤0.05). صحت روش‌های مبتنی بر LASSO در مقایسه با روش BGLR ۷ واحد افزایش داشت (P≤0.05). در مطالعه‌ای که با استفاده از داده شبیه‌سازی انجام شد متوسط صحت در روش‌های Bayes-A، B-LASSO gamma، B-LASSO beta و BGLR به ترتیب ۰/۸۷، ۰/۷۷، ۰/۷۷ و ۰/۷۴ گزارش گردید. که با نتایج پژوهش حاضر مطابقت داشت (Nicolazzi و همکاران، ۲۰۱۳).

در مطالعه‌ای گزارش گردید که برای صفات با اثرات QTL متوسط و بالا، روش Bayes-A نسبت به B-LASSO اثرات نشانگر SNP را با صحت بیشتر برآورد می‌کند (Cleveland و همکاران، ۲۰۱۰). در پژوهش کنونی با توجه به توزیع استفاده شده برای روش Bayes-A که بعضی از QTL ها متوسط و بزرگ

یک دلیل احتمالی این است که نشانگرها در LD کامل با ژنهای عامل<sup>۲۰</sup> نبوده و لذا واریانس ژنتیکی کل بطور کامل محاسبه نمی گردد (Guarini و همکاران، ۲۰۱۸). در مطالعه Nicolazzi و همکاران (۲۰۱۳) ضریب رگرسیون پیش‌بینی ژنومی برای روش‌های Bayes-A، B-LASSO gamma، B-LASSO، BGR و beta به ترتیب ۱/۰۰۹، ۰/۸۶۷، ۰/۸۵۹ و ۱/۳۹۵ گزارش داده شد که با نتایج پژوهش کنونی مطابقت دارد. مطابق جدول ۱، با افزایش وراثت‌پذیری در تمام روش‌های برآورد اثرات نشانگر میانگین مربعات خطای آزمایشی کاهش پیدا می‌کند. در وراثت‌پذیری ۰/۳ کمترین میزان خطا مربوط به روش Bayes-A (۱۲۱/۲) و بیشترین میزان مربوط به B-LASSO beta (۱۶۵/۲) بود. همچنین در وراثت‌پذیری ۰/۰۵ کمترین میزان خطای آزمایشی مربوط به روش Bayes-A (۱۳۱/۸) و بیشترین میزان را B-LASSO beta (۱۴۹/۸) نشان داد. در یک مطالعه که با داده‌های شبیه‌سازی انجام شد، با افزایش وراثت‌پذیری صفت، میانگین مربعات خطای آزمایشی در روش‌های پارامتری کاهش یافت که با نتایج به دست آمده در پژوهش کنونی مطابقت دارد (Howard و همکاران، ۲۰۱۴). در مطالعه محمدی و ستایی مختاری (۱۳۹۶) که با استفاده از داده‌های شبیه‌سازی شده انجام شده بود متوسط میانگین مربعات خطا در روش‌های Bیز A، B و LASSO و روش<sup>۱۱</sup> RKHS<sup>۱۱</sup> به ترتیب ۱/۸، ۱/۷، ۱/۸۶ و ۱/۹۵ گزارش داده شد. این نتایج نشان می‌دهد که شایستگی برازش مدل‌ها و متعاقب آن صحت پیش‌بینی توسط روش Bیز A نسبت به دیگر روش‌ها بهتر برآورد شد.

در جدول ۲ صحت پیش‌بینی ژنومی برای راهبردهای متفاوت پارامتر پیش‌تنظیمی  $\pi$  نشان داده شده است. با افزایش مقدار  $\pi$  تمام روش‌های برآورد اثر نشانگر، صحت افزایش یافت (جدول ۲). بین روش‌های مختلف برآورد اثرات نشانگر، در حالات متفاوت  $\pi$  از نظر صحت پیش‌بینی ژنومی اختلاف معنی‌داری دیده شد ( $P < 0.05$ ). بیشترین و کمترین مقدار صحت به ترتیب مربوط به Bayes-A در  $\pi$  برابر ۰/۵ (۰/۸۱) و برای روش BGLR در  $\pi$  برابر ۰/۱ (۰/۴۵) بود. Yang و همکاران (۲۰۱۵) بیان کردند

همکاران، ۲۰۰۵). استفاده از توزیع پیشین نمایی در روش بیزین LASSO منجر به درجه انقباض بیشتر اثرات نشانگر نسبت به روش Bayes-A می‌گردد (Van Raden و همکاران، ۲۰۰۹). در پژوهش کنونی صحت پیش‌بینی ژنومی در روش بیزین LASSO نسبت به روش Bayes-A در وراثت‌پذیری ۰/۳ مقدار ۲ واحد و در وراثت‌پذیری ۰/۰۵ مقدار ۵ واحد کاهش یافت که از لحاظ آماری معنی‌دار نبود. در یک مطالعه که از توزیع پیشین نمایی برای داده واقعی و شبیه‌سازی در گاو هلشتاین-فریزین استرالیایی استفاده شده بود، مقدار صحت بدست آمده توسط روش بیزین LASSO نسبت به روش Bayes-A برای داده واقعی متغیر چربی شیر مقدار ۴ واحد و برای داده شبیه‌سازی این مقدار ۱۰ واحد کمتر گزارش گردید (Nicolazzi و همکاران، ۲۰۱۳). Casella و Park (۲۰۰۸) پیشنهاد کردند که برای کاهش خطا در پارامترهای پیش‌تنظیمی باید تعداد داده‌ها را بزرگ انتخاب نمود. بالا بودن خطا و کم بودن صحت پیش‌بینی ژنومی در پژوهش کنونی را احتمالاً می‌توان به تعداد محدود داده‌های شبیه‌سازی شده نسبت داد. روش‌های بیزین LASSO برای نسل جدید تراشه‌های SNP برای تمام سطح ژنوم، احتمالاً در آینده روش مناسبی آماری برای برآورد صحت انتخاب باشند (Van Raden و همکاران، ۲۰۰۹).

با توجه به نتایج جدول ۱، ضریب رگرسیون به دست آمده برای تمام روش‌های برآورد اثرات نشانگر در وراثت‌پذیری ۰/۳ نسبت به وراثت‌پذیری ۰/۰۵ آریبی کمتری را نشان داد. مقدار ضریب رگرسیون در روش Bayes-A، B-LASSO gamma، B-LASSO beta و BGLR برای وراثت‌پذیری ۰/۳ به ترتیب ۱/۱۴، ۰/۹۳۵، ۰/۹۲۵ و ۱/۲۵ برآورد شد. با توجه به این مقادیر می‌توان گفت که روش Bayes-A نسبت به روش‌های دیگر ناریب‌تر است زیرا ضریب رگرسیون آن به عدد یک نزدیکتر می‌باشد. ضریب رگرسیون کمتر از یک نشان‌دهنده بیش‌برآورد<sup>۱۸</sup> پیش‌بینی ژنومی و بزرگتر از عدد یک نشانه کم‌برآورد<sup>۱۹</sup> پیش‌بینی ژنومی می‌باشد (Neves و همکاران، ۲۰۱۴). ضریب رگرسیون در روش‌های بیزین LASSO کوچکتر از یک برآورد گردید.

گزارش کردند که با افزایش میزان LD صحت پیش‌بینی ژنومی افزایش پیدا نمود.

که با افزایش میزان  $\pi$ ، صحت روش‌های آماری Bayes-A و Bayes-B افزایش می‌یابد که به طور کلی با نتایج به دست آمده در پژوهش کنونی مطابقت دارد. Chang و همکاران (۲۰۱۹)

جدول ۲- صحت برآورد اثرات نشانگر با راهبردهای متفاوت پارامتر پیش‌تنظیمی  $\pi$  تحت روش‌های مختلف برآورد اثرات نشانگر

پارامتر	راهبردها	Bayes-A	B-LASSO gamma	B-LASSO beta	BGLR
	۰/۱	۰/۶۹ <sup>a</sup>	۰/۶۶ <sup>ab</sup>	۰/۶۵ <sup>ab</sup>	۰/۴۵ <sup>b</sup>
	۰/۳	۰/۷۳ <sup>a</sup>	۰/۷۲ <sup>ab</sup>	۰/۷۱ <sup>ab</sup>	۰/۶۹ <sup>b</sup>
$\pi$	۰/۵	۰/۸۱ <sup>a</sup>	۰/۷۹ <sup>ab</sup>	۰/۷۹ <sup>ab</sup>	۰/۷۴ <sup>b</sup>

- حروف متفاوت در هر ردیف صحت انتخاب، نشان دهنده معنی داری بین روش‌ها را نشان می‌دهد (P<0.05).

در جدول ۳ صحت پیش‌بینی ژنومی برای راهبردهای متفاوت پارامتر پیش‌تنظیمی  $S^2$  نشان داده شده است. مطابق نتایج به دست آمده، با افزایش مقدار  $S^2$  در تمام روش‌های برآورد اثر نشانگر صحت کاهش یافت. از طرف دیگر در  $S^2$  های متفاوت بین روش‌های مختلف برآورد اثر نشانگر، اختلاف معنی داری از نظر صحت پیش‌بینی ژنومی بدست آمد (P<0.05). با حرکت از سمت ۰/۱ پارامتر پیش‌تنظیمی  $S^2$  به سمت راهبردهای ۱۰ و ۱۰۰ پارامتر، صحت پیش‌بینی ژنومی شیب کاهشی به خود گرفته و دلایل کاهش صحت احتمالاً بخاطر بیش برآورد<sup>۲۲</sup> زیاده‌تر صحت پیش‌بینی ژنومی است. در مطالعه‌ای نشان داده شد که با افزایش مقدار  $S^2$  مقدار صحت نیز کاهش می‌یابد که با نتایج پژوهش کنونی مطابقت دارد (Yang و همکاران، ۲۰۱۵). Nicolazzi و همکاران (۲۰۱۳) گزارش کردند که پارامترهای پیش‌تنظیمی تاثیر معنی داری روی صحت پیش‌بینی ژنومی روش‌های B-LASSO gamma و B-LASSO beta ندارند.



جدول ۳- صحت برآورد اثرات نشانگر با راهبردهای متفاوت پارامتر پیش تنظیمی  $S^2$  تحت روش های مختلف برآورد اثرات نشانگر

پارامتر	راهبردها	Bayes-A	B-LASSO gamma	B-LASSO beta	BGLR
$S^2$	۰/۰۱	۰/۶۹ <sup>a</sup>	۰/۶۶ <sup>ab</sup>	۰/۶۵ <sup>ab</sup>	۰/۴۵ <sup>b</sup>
	۰/۱	۰/۷۳ <sup>a</sup>	۰/۶۷ <sup>ab</sup>	۰/۷۱ <sup>ab</sup>	۰/۵۹ <sup>b</sup>
	۱۰	۰/۶۰ <sup>a</sup>	۰/۵۹ <sup>ab</sup>	۰/۵۳ <sup>ab</sup>	۰/۳۶ <sup>b</sup>
	۱۰۰	۰/۵۱ <sup>a</sup>	۰/۴۶ <sup>ab</sup>	۰/۴۲ <sup>ab</sup>	۰/۳۴ <sup>b</sup>

- حروف متفاوت در هر ردیف صحت انتخاب، نشان دهنده معنی داری بین روش ها را نشان می دهد (  $P < 0.05$  ).

آمدند. برای وراثت پذیری ۰/۰۵ مقدار پارامتر  $\pi$  برای روش های Bayes-A، B-LASSO gamma، B-LASSO beta و BGLR به ترتیب ۰/۶۷، ۰/۶۱، ۰/۵۶ و ۰/۷۳ و برای پارامتر  $S^2$  به ترتیب ۰/۰۴، ۰/۰۳، ۰/۰۲ و ۰/۰۵ بدست آمدند. با کاهش مقدار وراثت پذیری حد بهینه مقدار پارامتر پیش تنظیمی  $\pi$  برای روش های مختلف اثرات نشانگر افزایش و حد بهینه مقدار پارامتر  $S^2$  کاهش یافت.

در جدول ۴ مقدار بهینه پارامترهای پیش تنظیمی برای برآورد حداکثر صحت انتخاب ژنومی در وراثت پذیری های متفاوت تحت روش های مختلف برآورد اثرات نشانگر، نشان داده شده است. برای وراثت پذیری ۰/۳ برای روش های Bayes-A، B-LASSO gamma، B-LASSO beta و BGLR مقادیر بهینه پارامتر پیش تنظیمی  $\pi$  برای حصول حداکثر صحت پیش بینی ژنومی به ترتیب ۰/۵۳، ۰/۴۶، ۰/۴۱ و ۰/۵۶ و برای پارامتر  $S^2$  به ترتیب ۰/۱۱، ۰/۰۶، ۰/۰۵ و ۰/۱۲ بدست

جدول ۴- مقدار بهینه پارامترهای پیش تنظیمی برای برآورد حداکثر صحت انتخاب ژنومی تحت روش های مختلف برآورد اثرات نشانگر

وراثت پذیری	پارامترهای پیش- تنظیمی	Bayes-A	B-LASSO gamma	B-LASSO beta	BGLR
۰/۳	$\pi$	۰/۵۳	۰/۴۶	۰/۴۱	۰/۵۶
	$S^2$	۰/۱۱	۰/۰۶	۰/۰۵	۰/۱۲
۰/۰۵	صحت پیش بینی	۰/۸۹۳	۰/۸۶۷	۰/۸۵۹	۰/۷۹۶
	$\pi$	۰/۶۷	۰/۶۱	۰/۵۶	۰/۷۳
۰/۰۵	$S^2$	۰/۰۴	۰/۰۳	۰/۰۲	۰/۰۵
	صحت پیش بینی	۰/۶۹۴	۰/۶۵۴	۰/۶۸۳	۰/۶۰۳

## نتیجه گیری

نتایج حاصل از این مطالعه نشان می‌دهد که برای حداکثر نمودن صحت پیش‌بینی ژنومی مقدار بهینه پارامترهای پیش‌تنظیمی در معادلات مهم می‌باشد. لذا پیشنهاد می‌شود که برای صفات با وراثت‌پذیری بالا مقادیر بهینه پارامتر پیش‌تنظیمی  $\pi$  و  $S^2$  برای روش‌های بیزین به ترتیب  $0/41$  تا  $0/53$  و  $0/05$  تا  $0/11$  و این مقادیر برای روش BGLR،  $0/56$  و  $0/12$  پیشنهاد می‌گردد. برای صفات با وراثت‌پذیری پایین مقادیر بهینه پارامتر پیش‌تنظیمی  $\pi$  و  $S^2$  برای روش‌های بیزین به ترتیب  $0/56$  تا  $0/67$  و  $0/02$  تا  $0/04$  و این مقادیر برای روش BGLR،  $0/73$  و  $0/05$  پیشنهاد می‌گردد.

## منابع

- محمدی، ی. و ستایی مختاری، م. (۱۳۹۶). روش‌های پارامتری و ناپارامتری برای بررسی صحت انتخاب ژنومی با معماری‌های ژنتیکی افزایشی و غالبیت. پژوهش‌های تولیدات دامی. شماره ۱۸. صص ۱۶۷-۱۶۱.
- Chang, L.Y., Toghian, S., Aggrey, S.E. and Rekaya R. (2019). Increasing accuracy of genomic selection in presence of high density marker panels through the prioritization of relevant polymorphisms. *BMC Genetics*. 20: 21.
- De los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D. and Calus M.P.L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 193: 327-345.
- De los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K., and Cotes, J.M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*. 182:375-385.
- Harris, B.L., Johnson, D.L., and Spelman, R.J. (2008). Genomic selection in New Zealand and the implications for national genetic evaluation. Proc. 36th ICAR Biennial Session, Niagara Falls, Canada, 38:325-330.
- Howard, R., A. L. Carriquiry and W. D. Beavis. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3*. 4: 1027-1046.
- Gaspa, G., Pintus, M.A., Nicolazzi, E.L., Vicario, D., Valentini, C., and Macciotta, N.P.P. (2013). Use of principal component approach to predict direct genomic breeding values for beef traits in Italian Simmental cattle. *Journal Animal Science*. 91:29-37.
- Guarini, A.R., Lourenco, D.A.L., Brito, L.F., Sargolzaei, M., Baes, C., Miglior, F., Misztal, I. and Schenkel, F.S. (2018). Comparison of genomic predictions for lowly heritable traits using multi-step and single-step genomic best linear unbiased predictor in Holstein cattle. *Journal of Dairy Science*. 101:1-11.
- Hayes, B., Bowman, P., Chamberlain, A. and Goddard, M. (2009). Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*. 92:433-443.
- Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157: 1819-1829.
- Neves, H.H.R., Carneiro, R., O'Brien, A. M., Utsunomiya, Y.T., do Carmo, A.S., Schenkel, F.S., Sölkner, J., McEwan, J.C., Van Tassell, C.P., Cole, J.B., da Silva, M.V., and Queiroz, S. A. (2014). Accuracy of genomic predictions in *Bos indicus* (Nelore) cattle. *Genetics Selection Evolution*. 46:17.
- Nicolazzi, E., Negrini, R., Amadori, J., Chamberlain, Goddard, M.E., Marsan, P.A., and Hayes B.J. (2013). Effect of Prior Distributions on Accuracy of Genomic Breeding Values for Two Dairy Traits. *Italian Journal of Animal Science*. 12: 551-561.
- Park, T., and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*. 103: 681-686.

- Pintus, M.A., Nicolazzi, E.L., Van Kaam, J.B., Biffani, S., Stella, A., Gaspa, G., Dimauro, C., and Macciotta, N.P. (2013). Use of different statistical models to predict direct genomic values for productive and functional traits in Italian Holstein. *Journal Animal Breeding and Genetics*. 130:32-40.
- ter Braak, C.J.F., Boer, M.P., and Bink, M.C.A.M. (2005). Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. *Genetics*. 170:1435-1438.
- Sargolzaei, M., and Schenkel, F.S. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics*. 25: 680-1.
- Schefers, M.J. and Weigel, K.A. (2020). Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Animal Frontiers*. 2(1): 24-37.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal Dairy Science*. 92:16-24.
- Verbyla, K., Bowman, P.J., Hayes, B.J., and Goddard, M.E. (2010). Sensitivity of genomic selection to using different prior distributions. *BMC Proceedings* 4(Suppl.1):S5.
- Yang, w., Chen, C., and Tempelman, R.J. (2015). Improving the computational efficiency of fully Bayes inference and assessing the effect of misspecification of hyperparameters in whole-genome prediction models. *Genetics Selection Evolution*. 47: 13.
- Yi, N., and Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics*. 179:1045-1055.
- Zhang, H., Yin, L., Wang, M., Yuan, X., and Liu X. (2019). Factors Affecting the Accuracy of Genomic Selection for Agricultural Economic Traits in Maize, Cattle, and Pig Populations. *Frontiers in Genetics*. 10: 180-189.

