تولیدات دامی
دوره ۱۸ ■ شماره ۱ ■ بهار ۱۳۹۵
صفحه‌های ۱۱-۱

# مقایسهٔ دقت برآورد ارزش اصلاحی ژنومی برای صفات تولید در گاوهای هلشتاین ایران با روش‌های پارامتری و ناپارامتری

یحیی محمدی[۱]، محمدمهدی شریعتی[۲]*، سعید زره‌داران[۳]، محمد رزم‌کبیر[۴]، محمدباقر صیادنژاد[۵]، محمدباقر زندی[۶]

۱. دانشجوی دکتری، گروه علوم دامی، دانشکدهٔ کشاورزی، دانشگاه فردوسی مشهد، ایران
۲. استادیار، گروه علوم دامی، دانشکدهٔ کشاورزی، دانشگاه فردوسی مشهد، ایران
۳. استاد، گروه علوم دامی، دانشکدهٔ کشاورزی، دانشگاه فردوسی مشهد، ایران
۴. استادیار، گروه علوم دامی، دانشکدهٔ کشاورزی، دانشگاه کردستان، ایران
۵. کارشناس ارشد، مرکز اصلاح نژاد و بهبود تولیدات دامی کشور، ایران
۶. کارشناس ارشد، بخش پرورش و اصلاح اسب، مرکز اصلاح نژاد و بهبود تولیدات دامی کشور، ایران

تاریخ وصول مقاله: ۱۳۹۴/۰۵/۳۱ تاریخ پذیرش مقاله: ۱۳۹۴/۰۷/۱۵

## چکیده

انتخاب ژنومی ابزاری جدید برای برآورد ارزش‌های اصلاحی صفات کمّی با استفاده از نشانگرهای مولکولی است. معیار کارایی انتخاب ژنومی، دقت پیش‌بینی ارزش‌های اصلاحی ژنومی است. در تحقیق حاضر، دقت پیش‌بینی ژنومی روش‌های پارامتری و ناپارامتری در ۳۴۵ گاو هلشتاین محاسبه شد. صفات مطالعه‌شده مقدار شیر، میزان چربی، میزان پروتئین، و سلول‌های سوماتیک شیر بود. دو روش پارامتری بهترین پیش‌بینی نااریب خطی ژنومیک (GBLUP) و بیز B و دو روش ناپارامتری فضای هیلبرت با هستهٔ بازآفرین (RKHS) و شبکه‌های عصبی (NN) برای برآورد تأثیرات نشانگرها و پیش‌بینی و دقت ارزش‌های اصلاحی ژنومی استفاده شد. دقت پیش‌بینی ژنومی در دامنهٔ ۰/۳۹ (برای سلول‌های سوماتیک) تا ۰/۷۳ (برای تولید چربی) محاسبه شد. بیز B و دو روش ناپارامتری در مقایسه با روش GBLUP دچار کوچک‌کردن بیشتر پیش‌بینی‌ها شده بودند (منحنی ضریب رگرسیون ارزش اصلاحی ژنومی بر ارزش اصلاحی کلاسیک بیشتر از یک به‌دست آمد). درمقایسه با همهٔ روش‌ها دقت بیز B بیشتر شد. میانگین مربعات خطای پیش‌بینی برای روش GBLUP به نسبت دیگر روش‌ها برای صفات مطالعه‌شده کمتر برآورد شد. مدل رگرسیون بیز B برای انتخاب ژنومی در این جمعیت مطلوب بود، ولی برای بهبود دقت با این روش نیاز به کاهش آب‌رفتگی پیش‌بینی در این روش است.

**کلیدواژه‌ها:** انتخاب ژنومی، دقت انتخاب، روش‌های پارامتری و ناپارامتری، نشانگرهای مولکولی.

* نویسنده مسئول
Email: mm.shariati@um.ac.ir

**مقدمه**

در روش‌های مرسوم اصلاح دام انتخاب برای صفات مهم اقتصادی براساس رکوردهای فنوتیپی افراد و خویشاوندان آنها صورت می‌پذیرد، بنابراین ارزش‌های اصلاحی افراد به شیوهٔ بهترین پیش‌بینی نااریب خطی (BLUP) پیش‌بینی می‌شود. در این روش، از اطلاعات ژنومی برای کمک به انتخاب حیوان استفاده نمی‌شود و حیوانات براساس مجموع اثر جایگزینی ژن‌ها که ارزش اصلاحی نامیده می‌شود، انتخاب می‌شوند. پیشرفت ژنتیکی حاصل از انتخاب با این روش برای مقداری از صفات موفق است، ولی برای صفاتی که محدود به جنس‌اند و یا اینکه در مراحل پایانی زندگی حیوان یا حتی بعد از کشتار دام اندازه‌گیری می‌شوند، این روش معایبی دارد [۲۱]. با پیشرفت استفاده از ژنتیک مولکولی در سال ۱۹۶۰، استفاده از اطلاعات نشانگرها در برنامه‌های اصلاح نژاد شروع شد و تئوری انتخاب براساس نشانگر به‌طور گسترده بسط و توسعه یافت [۲]. با اینکه تحقیقات وسیعی در زمینهٔ انتخاب به کمک مارکر (Marker assisted selection) انجام گرفت، کاربردهای آن محدود و افزایش پیشرفت ژنتیکی به کمک آن کم بوده است. مطالعات اخیر با استفاده از تراشه‌های SNPها (Single nucleotide polymorphisms) در گاو نشان داد که صفات کمّی اغلب تحت‌تأثیر تعداد زیادی ژن با اثر کم هستند. همچنین به استثنای تعداد محدودی ژن، بیشتر ژن‌های مؤثر بر صفات کمّی به مقدار کم در این واریانس ژنتیکی صفت نقش دارند زیرا بیشتر صفات اقتصادی تحت‌تأثیر ژن‌های زیادی قرار دارند با ردیابی تعداد کمی از آنها به‌وسیلهٔ نشانگرهای ژنتیکی، فقط قسمت کمی از واریانس ژنتیکی توجیه می‌شود. به‌علاوه، اثر تک‌ژن‌ها بسیار کم است و داده‌های زیادی برای برآورد اثر دقیق آنها لازم است [۱۶].

برای رفع این مشکلات و باتوجه به کشف تعداد بسیار زیاد نشانگر SNP و پیشرفت در فناوری ژنوتیپ‌کردن با هزینهٔ کم و دقت زیاد، این امکان فراهم شد که انتخاب به کمک نشانگر در مقیاس ژنوم را انجام داد که به این روش اصطلاحاً انتخاب ژنومی گفته می‌شود [۲۰]. انتخاب ژنومی بر این فرض استوار است که همهٔ QTLهای مؤثر بر صفات مورد نظر در حالت عدم تعادل پیوستگی با حداقل یک SNP است و بنابراین ارزش اصلاحی حیوان با ارزیابی همزمان تعداد زیادی از نشانگرها در سطح کل ژنوم و حاصل جمع تأثیرات آنها برآورد می‌شود [۱۶]. برای افزایش دقت پیش‌بینی ژنومی، تعداد نشانگر کافی در سطح ژنوم، اطلاعات فنوتیپی، و مدل‌های آماری مناسب برای برآورد تأثیرات نشانگرها لازم است [۱۴].

در انتخاب ژنومی به‌طور معمول تعداد متغیرهای پیش‌بینی‌شده (نشانگرها) بسیار بیشتر از تعداد مشاهدات (فنوتیپ‌ها) است. بنابراین برآورد تأثیرات نشانگرها با روش‌های رگرسیونی ساده امکان‌پذیر نیست [۱ و ۲۲]. برای رفع این مشکل استفاده از روش‌های جریمه‌ای همراه با انتخاب متغیر همچون Bayes B، Bayes Cπ، و Bayes L (بیز LASSO)، و روش‌های برآورد کوچک‌کردن (Shrinkage Estimation) مانند بهترین پیش‌بینی نااریب خطی ژنومی یا GBLUP، رگرسیون ریدج یا RRBLUP، و Bayes A و همچنین روش‌های نیمه‌پارامتری و ناپارامتری چون روش RKHS و NN امکان‌پذیر است [۲۲].

نتایج مطالعات انتخاب ژنومی با داده‌های واقعی، نشان می‌دهد که برای صفتی مانند چربی شیر که تحت‌تأثیر ژن‌های بزرگ‌اثر (DGAT1) است، نتایج روش‌های بیز Aو B بهتر از روش‌های GBLUP است. ولی برای صفتی مانند پروتئین شیر که اکثر واریانس صفت با ژن‌هایی با اثر جزئی کنترل می‌شود و معماری ژنتیکی آن از مدل بی‌نهایت جایگاه (Infinitesimal model) تبعیت می‌کند، عملکرد روش‌های بیزی و GBLUP تقریباً مشابه است [۷].

روش‌های پارامتری براساس مدل‌های خطی است و به‌آسانی قادر به برازش داده‌ها داخل مدل بدون اثر بیش‌برآورد هستند. با این حال، ارتباط بین ارزش‌های اصلاحی و نشانگرهای ژنتیکی احتمالاً پیچیده‌تر از یک ارتباط خطی ساده (به‌ویژه وقتی که تعداد نشانگرهای SNP خیلی زیاد هستند) است [۱۵]. علاوه بر تأثیرات افزایشی، تأثیرات غیرافزایشی نیز سهمی را در واریانس ژنتیکی صفات ایفا می‌کنند. به‌طورکلی، در روش‌های ناپارامتری فرضیات در مقایسه با مدل‌های پارامتری ساده‌تر است و بدون نیاز به مدل‌های پیچیده، تأثیرات غیرافزایشی برآورد می‌شود.

در تحقیق حاضر، مقایسهٔ چهار روش آماری GBLUP، Bayes B، RKHS، و NN در پیش‌بینی ارزش‌های اصلاحی ژنومی و دقت آنها برای صفات تولید (مقدار شیر، میزان چربی، میزان پروتئین، و سلول‌های سوماتیک شیر) در گاوهای هلشتاین ایران مقایسه می‌شود.

## مواد و روش‌ها

این تحقیق با همکاری مرکز اصلاح نژاد و بهبود تولیدات دامی کشور در زمینهٔ گاوهای هلشتاین ایران انجام شد. ژنوتیپ تعداد ۳۴۵ گاو نر که در این مرکز دارای سابقه و رکورد تولید از خود، دختران یا والدین آنها بودند، تعیین شد. در این زمینه، ژنوتیپ حیوانات با تراشهٔ نشانگری ۵۰ کیلوباز (شرکت ایلومینا سان دیاگو، کانادا) تعیین شد. کنترل کیفیت به کمک فرایند تکرار معیار انتخاب SNP، نشانگرهای با فراوانی آلل مینور بالاتر از ۰/۰۲ (MAF<۰/۰۲)، میزان فراخوانی آللی بیشتر از ۰/۹۸، و تست هاردی واینبرگ (HWE) بیشتر از ۰/۰۰۰۰۱ انتخاب شدند. نشانگرهای انتخابی در مرحلهٔ بعد برای برآورد عدم تعادل پیوستگی (LD) استفاده شدند. جدول ۱ تعداد نشانگرهای باقیمانده بعد از فاکتورهای تصحیح را نشان می‌دهد که از تعداد ۵۴۰۰۱ نشانگر تعداد ۴۲۳۵۴ نشانگر باقی ماندند.

ارزش‌های اصلاحی کلاسیک (EBV)، وراثت‌پذیری، و مؤلفه‌های واریانس برای صفات بررسی‌شده با روش BLUP به کمک نرم‌افزار DMU و با رابطهٔ ۱ برآورد شد. در تحقیق حاضر، داده‌ها به کمک آنالیز تک‌صفتی برای رکوردهای روزآزمون استفاده شدند. این رکوردها با مدل رگرسیون ثابت که نوعی مدل تکرارپذیری است، آنالیز شدند. تمام گاوهای ژنوتیپ‌شده در این تحقیق دارای رکورد تولیدی بودند و ارزش‌های اصلاحی کلاسیک این حیوانات به کمک رکوردهای خود یا خویشاوندان (دختران یا والدین) برآورد شد، اما تعدادی از این گاوها دارای شمارهٔ انجمن ملی پرورش‌دهندگان دام نبودند که برای برآورد ارزش اصلاحی کلاسیک نیازی به این شماره نبود.

$$y = 1\mu + Za + e \qquad (۱)$$

در این رابطه: y بردار رکوردهای روزآزمون، μ میانگین فنوتیپی، a بردار تأثیرات ژنتیکی افزایشی، e بردار تأثیرات تصادفی باقیمانده، Z ماتریس ضرایب a، و 1 نیز بردار واحد است.

**جدول ۱. تعداد نشانگرهای حذف‌شده از روش کنترل کیفیت**

| معیارهای انتخاب | نرخ معیار انتخاب | تعداد نشانگرهای خروجی |
|---|---|---|
| فراوانی آلل مینور | ۰/۰۲ | ۷۰۳۲ |
| فراخوانی آلل | ۰/۹۸ | ۳۷۱۱ |
| تست هاردی واینبرگ | ۰/۰۰۰۰۱ | ۹۰۴ |

**جدول ۲. آماره‌های توصیفی و قابلیت اعتماد ارزش اصلاحی کلاسیک برای صفات بررسی‌شده**

| صفات | تعداد رکوردها | میانگین | انحراف معیار | حداقل | حداکثر | وراثت‌پذیری صفت | قابلیت اعتماد |
|---|---|---|---|---|---|---|---|
| مقدار شیر | ۸۵۶۴۳۲ | ۲۸/۱۱ | ۷/۶۷ | ۴/۵ | ۵۲/۵ | ۰/۲۰۶ | ۰/۶۹ |
| میزان چربی | ۷۶۵۳۲۱ | ۰/۹۲ | ۰/۳۳ | ۰/۰۴ | ۳/۶۵ | ۰/۲۱۱ | ۰/۶۳ |
| میزان پروتئین | ۵۹۱۲۳۴ | ۰/۹۳ | ۰/۲۶ | ۰/۰۳ | ۲/۸۹ | ۰/۲۲۵ | ۰/۶۱ |
| سلول‌های سوماتیک شیر | ۴۳۵۶۷۸ | ۳۴۰۲۱۱ | ۸۶۷۵۴ | ۸۷۶۱۱ | ۵۵۱۱۷۸ | ۰/۰۷۲ | ۰/۵۹ |

ارزش اصلاحی کلاسیک به‌عنوان متغیر پاسخ برای برآورد تأثیرات SNP و دقت پیش‌بینی‌های ژنومی استفاده شد. برای صفات مقدار شیر، میزان چربی، و میزان پروتئین برآورد وراثت‌پذیری متوسط و برای صفت سلول‌های سوماتیک شیر وراثت‌پذیری کم و نیز قابلیت اعتماد ارزش‌های اصلاحی صفات مطالعه‌شده نیز زیاد بود (جدول ۲).

در این تحقیق، کل حیوانات ژنوتیپ‌شده براساس سال تولد به دو گروه جمعیت مرجع و جمعیت تأیید تقسیم‌بندی شدند. حیوانات متولد سال‌های ۱۳۸۳ تا ۱۳۸۸ به‌عنوان جمعیت مرجع و حیوانات متولد سال‌های ۱۳۸۹ تا ۱۳۹۱ به‌عنوان جمعیت تأیید در نظر گرفته شدند. تعداد افراد جمعیت مرجع و تأیید به‌ترتیب ۲۵۵ و ۹۰ حیوان بودند. از چهار روش آماری متفاوت برای پیش‌بینی ارزش اصلاحی ژنومی به این شرح استفاده شد:

**روش بهترین پیش‌بینی نااریب خطی ژنومی (GBLUP)**

در روش GBLUP، از رابطهٔ ۲ برای پیش‌بینی ارزش اصلاحی ژنومی استفاده شد:

$$y = 1_n \mu + Zg + e \qquad (۲)$$

در این رابطه: y بردار ارزش‌های فنوتیپی، μ میانگین فنوتیپی، g بردار ارزش‌های اصلاحی ژنومی، e بردار تأثیرات تصادفی باقیمانده، Z ماتریس ضرایب که مشاهدات را به بردارهای ارزش‌های اصلاحی مرتبط می‌کند، و $1_n$ نیز بردار واحد است. در این مدل اثر نشانگرها به‌صورت تصادفی با واریانس یکسان در نظر گرفته شده است. در این مدل فرض می‌شود که $g \sim N(0, G\sigma_g^2)$ و $e \sim N(0, R\sigma_e^2)$ و R، ماتریس قطری، $\sigma_e^2$ واریانس باقیمانده، $\sigma_g^2$ واریانس ژنتیکی افزایشی، و G ماتریس روابط خویشاوندی ژنومی که با رابطهٔ ۳ محاسبه شد [۲۵]:

$$Gj_{j'} = \frac{1}{m}\sum_{i=1}^{m}\frac{(X_{ij} - 2p_i)(X_{ij'} - 2p_i)}{2p_i(1-p_i)} \qquad (۳)$$

در این رابطه: $X_{ij}$ تعداد کپی آلل مرجع برای iامین SNP در jامین فرد، $p_i$ نیز فراوانی آلل مرجع، و m تعداد نشانگر است.

**روش بیز B**

در این روش فرض بر این است که فقط تعدادی از جایگاه‌های ژنی، کل واریانس ژنتیکی را توجیه می‌کنند و بسیاری از جایگاه‌های ژنی در واریانس ژنتیکی نقش ندارند. رابطهٔ ۴ مدل آماری بیز B را نشان می‌دهد:

$$y_j = \mu + \sum_{i=1}^{m} x_{ij}\beta_i\delta_i + e_j \qquad (۴)$$

در این رابطه: y فنوتیپ حیوان j، μ میانگین، m تعداد جایگاه نشانگر، x ژنوتیپ نشانگر در جایگاه i (jاُمین آلل) است که به‌صورت صفر، یک، و دو یعنی تعداد آلل SNP موجود در حیوان j کدگذاری می‌شوند. $\beta i$ نیز اثر

جایگزینی آللی در جایگاه i و $\delta_i$ متغیر اختصاصی این مدل و نشان عدم حضور (با احتمال π) یا حضور (با احتمال 1-π) جایگاه i در مدل است.

$$\delta_{i=1\Rightarrow\beta_i} \sim N(0,\sigma^2_i)$$
$$\delta_{i=0\Rightarrow\beta_i} = 0$$

فرض اصلی این روش این‌است که بسیاری از SNPها در محل‌هایی از ژنوم قرار دارند که فاقد QTL است و بر صفت تأثیری ندارند، در حالی‌که فقط بخش اندکی از SNPها در حالت LD با QTLاند و درنتیجه دارای اثر هستند. به‌طورکلی، π نشانهٔ نسبت قابل انتظار از SNPهاست که در حالت عدم پیوستگی با QTLها در مقایسه با تعداد کل SNPها است. تأثیرات SNP از توزیع t نمونه‌گیری شده است ولی واریانس تأثیرات با احتمال π از تودهٔ صفر و با احتمال 1-π از توزیع مربع کای معکوس-مقیاس‌شده نمونه‌گیری می‌شوند [۲۰]. در روش بیز B مشکل مشخص‌بودن مقدار π است، در صورتی‌که مقداری که انتخاب می‌شود با توزیع حقیقی تأثیرات SNP ناسازگار و متناقض باشد به‌طور معنی‌داری بر میزان دقت تأثیرات برآوردشده تأثیر منفی دارد. برای حل این مشکل، از روش نمونه‌گیری که توزیع پیشین یکنواخت (π = uniform (0,1)) دارد و در کنار تمام پارامترهای دیگر در تکرارهای MCMC نمونه‌گیری می‌شود، استفاده شد. زمانی‌که یک همگرایی حاصل شد، پارامتر برابر میانگین توزیع پسین خود آن در نظر گرفته می‌شود و الگوریتم‌ها مجدداً برای برآورد تأثیرات SNP استفاده می‌شوند.

**فضای هیلبرت با هستهٔ بازآفرین (RKHS)**

این روش به‌عنوان روش نیمه‌پارامتری برای برآورد ارزش اصلاحی ژنومی معرفی شد [۱۵].

مدل آماری این روش به‌صورت رابطهٔ ۵ است:

$$y = 1\mu + K_h\alpha + e \qquad (۵)$$

در این رابطه: y بردار فنوتیپ‌ها، μ ثابت نامشخص، 1 بردار یک، $K_h$ ماتریس مثبت-معین کرنل، α و e با فرض توزیع پیشین نرمال برابر $\alpha \sim N(0, K_h\sigma^2_a)$ و $e \sim N(0, I\sigma^2_e)$ است. ورودی‌های ماتریس $K_h$ کرنل به‌صورت رابطهٔ ۶ تعریف می‌شود:

$$K_h(x_i, x_j) = \exp(-hd_{ij}), \qquad (۶)$$

در این رابطه: $d_{ij}$ مربع فاصلهٔ اقلیدسی محاسبه‌شده بین افراد i و j براساس ژنوتیپ نشانگرهای SNP آنها و h پارامتر هموارسازی به‌صورت $h = 2/d^*$ است که $d^*$ میانگین $d_{ij}$ است. انتخاب یک کرنل مهمترین مرحله در استفاده از الگوریتم‌های بر پایهٔ کرنل است [۱۴]. در این تحقیق از کرنل‌نمایی استفاده شد.

**شبکه‌های عصبی (NN)**

شبکهٔ عصبی مصنوعی، مجموعه‌ای از نرون‌های به هم متصل در لایه‌های گوناگون است که اطلاعاتی را برای یکدیگر ارسال می‌کنند. ساده‌ترین شکل شبکه فقط دو لایه دارد: لایهٔ ورودی و لایهٔ خروجی. شبکه شبیه سیستم ورودی-خروجی عمل می‌کند و ارزش نورون‌های ورودی را برای محاسبهٔ ارزش نرون خروجی استفاده می‌کند. ارتباط میان یک ورودی و خروجی به‌وسیلهٔ ضریب α که نمایندهٔ اهمیت نسبی ورودی مذکور در محاسبهٔ ارزش خروجی است، مشخص می‌شود. به این ترتیب، ارزش نورون خروجی مشاهدهٔ t از رابطهٔ ۷ محاسبه می‌شود:

$$Net_t = \alpha_0 x_{0t} + \alpha_1 x_{1t} + \alpha_2 x_{2t} = \sum_{i=0}^{2} \alpha_i x_{it} \qquad (۷)$$

سپس نورون خارجی ارزش حاصل را با استفاده از یک تابع تبدیل یا فعال‌سازی (محرک) پردازش می‌کند. در این تحقیق، نورون‌های ورودی همان متغیرهای مستقل (نشانگرها) و نورون‌های خروجی همان برآورد متغیر

وابسته (تأثیرات نشانگر) هستند. به‌طورکلی، نقش نرون‌ها در شبکه‌های عصبی پردازش اطلاعات است و این امر در شبکه‌های عصبی مصنوعی به‌وسیلهٔ یک پردازشگر ریاضی که همان تابع فعال‌سازی است، انجام می‌شود. تابع فعال‌سازی می‌تواند خطی یا غیرخطی باشد. تابع فعال‌سازی، براساس نیاز خاص مسئله‌ای که قرار است با شبکهٔ عصبی حل شود، از سوی طراح انتخاب می‌شود. برای بهره‌برداری از توانایی شبکه‌های عصبی باید از توابع فعال‌سازی غیرخطی استفاده کنند. این امر اجازه می‌دهد که شبکه الگوهای غیرخطی مناسبی از مجموعه داده‌های پیچیده تولید کند. در این تحقیق از تابع نمایی و ۵ لایهٔ پنهان و ۲ نورون استفاده شد.

از نرم‌افزار BGLR [۵] به‌منظور اجرای روش‌های GBLUP، بیز B، و روش RKHS و برای اجرای روش NN از نرم‌افزار brnn [۲۴] در نرم‌افزار R استفاده شد. برای محاسبهٔ تأثیرات نشانگر در تمام مدل‌های بیزی از الگوریتم MCMC استفاده شد. بدین‌منظور از یک زنجیره با طول ۵۰۰۰۰ تکرار استفاده شد. ۲۰۰۰۰ نمونهٔ اول به‌عنوان دورهٔ گرم‌کردن از نتایج خروجی و ۳۰۰۰۰ نمونهٔ باقیمانده برای برآورد میانگین توزیع پسین هر پارامتر استفاده شد. برای اطمینان از همگرایی نمونه‌گیری گیبس، از خودهمبستگی بین نمونه‌ها و رسم پلات‌ها استفاده شد.

مطالعات متعددی روی دقت انتخاب ژنومی صورت گرفته است. دقت ارزیابی ژنومی، از همبستگی ارزش اصلاحی واقعی ناشناخته و ارزش اصلاحی ژنومی محاسبه می‌شود. چون TBV نامعلوم است به کمک اطلاعات دردسترس دقت برآورد را محاسبه می‌کنند. از رابطهٔ ۸ برای برآورد دقت انتخاب ژنومی استفاده می‌شود [۴]:

$$r_{(GEBV,TBV)} = \frac{cor(GEBV, EBV)}{cor(EBV, TBV)} = \frac{cor(GEBV, EBV)}{\sqrt{r^2_{EBV}}} \qquad (۸)$$

در این رابطه: $r^2_{EBV}$ قابلیت اعتماد ارزش اصلاحی کلاسیک است. برای برآورد دقت انتخاب ژنومی از ضریب همبستگی پیرسن بین GEBV و EBV (r(GEBV, EBV)) تقسیم بر جذر ضریب وراثت‌پذیری استفاده شده است [۴ و ۷]. برای محاسبهٔ دقت انتخاب ژنومی در جمعیت تأیید، ابتدا تأثیرات نشانگر در جمعیت مرجع برآورد شد و سپس این اثر برای محاسبهٔ ارزش اصلاحی ژنومی در جمعیت تأیید استفاده شد. دقت برآورد ارزش‌های اصلاحی ژنومی توسط معادلهٔ ۸ برآورد شد.

میانگین مربعات خطای پیش‌بینی بین ارزش اصلاحی ژنومی پیش‌بینی‌شده و ارزش اصلاحی کلاسیک در داده‌های جمعیت مرجع به‌عنوان اندازه‌ای از سازگاری مدل به داده‌ها است. هرقدر میانگین مربعات خطای پیش‌بینی بیشتر باشد، نشانهٔ دقت کمتر پیش‌بینی‌های ژنومی و کم‌بودن این عدد نشانهٔ زیادبودن دقت پیش‌بینی‌هاست. ضریب رگرسیون ارزش اصلاحی کلاسیک بر ارزش اصلاحی ژنومی در جمعیت مرجع، اندازه‌ای از برآورد بیش از حد (Inflation) یا برآورد کمتر از حد (deflation) از پیش‌بینی ژنومی را نشان می‌دهد. انتظار این است که این ضریب نزدیک به یک باشد که در این حالت نشانهٔ این است که مقیاس ارزش اصلاحی ژنومی مشابه ارزش اصلاحی کلاسیک است و هیچ‌گونه برآورد بیش از حد یا کمتر از حد ندارد.

## نتایج و بحث

بعد از کنترل کیفی داده‌های ژنوتیپی، میانگین (انحراف معیار) فراوانی آلل‌های مینور ۰/۲۰۱ (۰/۱۱۲) و میانگین

(میانه) عدم تعادل پیوستگی ($r^2$) بین جفت نشانگرهای کناری ۰/۲۶۵ (۰/۱۴۲) برآورد شد.

دامنهٔ دقت پیش‌بینی ژنومی ($r_{GBV,EBV}$) در صفات مورد نظر از ۰/۳۹ برای سلول‌های سوماتیک تا ۰/۷۳ برای تولید چربی متغیر بود. متوسط دقت پیش‌بینی ژنومی در صفات برای روش‌های GBLUP، Bayes B، RKHS، و NN به‌ترتیب ۰/۵۷، ۰/۶۲، ۰/۴۷، و ۰/۴۶ برآورد شد(جدول ۳).

متوسط دقت پیش‌بینی ژنومی در داده‌های این تحقیق به‌طور کلی کمتر از داده‌های شبیه‌سازی است. در یک تحقیق، در داده‌های شبیه‌سازی متوسط دقت ژنومی برای روش‌های GBLUP، Bayes A، و Bayes B به ترتیب ۰/۷۳۲، ۰/۷۹۸، و ۰/۸۴۸ حاصل شد [۲۰]. از دلایل کم‌بودن دقت در داده‌های واقعی، احتمالاً افزایش معنی‌دار تعداد QTL مؤثر در آنهاست که در مطالعات نقشه‌یابی QTL در نظر گرفته می‌شود [۱۷ و ۲۰]. از طرف دیگر، در دادهٔ شبیه‌سازی در مقایسه با دادهٔ واقعی مقدار LD و تعداد نشانگرها مشخص و معمولاً بیشتر است. یکی دیگر از دلایل کم‌بودن دقت می‌تواند کم‌بودن تعداد نمونه باشد.

در مطالعات با دادهٔ واقعی معماری ژنتیکی صفات تا حدود زیادی ناشناخته‌اند و گروه‌بندی صفات براساس معماری آنها آسان نیست. در گاو شیری تعدادی صفات تحت‌تأثیر ژن‌های بزرگ‌اثر است، به‌طوری که این ژن‌ها نسبت زیادی از واریانس ژنتیکی صفت را توجیه می‌کنند. یکی از این صفات چربی شیر است که تحت‌تأثیر ژن بزرگ‌اثر DGAT1 است که اثر زیادی بر درصد چربی شیر گاو دارد. در توصیف این ویژگی‌ها نتایج چند تحقیق گزارش شده است که مدل‌هایی با توزیع پیشین انتهای کشیده متأثر از نشانگر نظیر بیز A و روش‌های انتخاب متغیر نظیر بیز Cπ و بیز LASSO با دقت بیشتری در مقایسه با روش GBLUP تأثیرات نشانگر را برآورد می‌کنند. دقت برآورد ارزش ژنومی صفت چربی شیر با روش بیز B (۰/۷۳) و روش GBLUP (۰/۵۹) محاسبه شد (جدول ۳). این نتایج در مطابقت با نتایج سایر مطالعات بود [۸ و ۱۲]. به‌طورکلی، دقت برآورد ارزش‌های اصلاحی روش‌های پارامتری (G- BLUP و بیز B) به نسبت روش‌های ناپارامتری بیشتر است. مطابق با نتایج این تحقیق، روش SNPBLUP در مقایسه با روش RKHS، ۰/۱ ارزش اصلاحی ژنومی را با دقت بیشتر برآورد کرد [۱۸]، اما در تحقیقی دیگر برای دقت برآورد ارزش ژنومی بین روش NN و روش بیزی، صفات نمرهٔ ماربلینگ در گاو انگوس اختلاف معنی‌داری بین دو روش نداشت [۲۳].

**جدول ۳. دقت پیش‌بینی ژنومی از صفات تولیدی در گاوهای هلشتاین ایران با مدل‌های متفاوت آماری**

| صفات | $r_{GEBV,EBV}$ | | | |
|---|---|---|---|---|
| | GBLUP | Bayes B | RKHS | NN |
| تولید شیر | ۰/۶۴ | ۰/۶۸ | ۰/۴۹ | ۰/۴۷ |
| تولید چربی | ۰/۵۹ | ۰/۷۳ | ۰/۵۲ | ۰/۵۱ |
| تولید پروتئین | ۰/۵۷ | ۰/۵۸ | ۰/۴۸ | ۰/۴۶ |
| شمارش سلول‌های سوماتیک | ۰/۴۹ | ۰/۴۹ | ۰/۴۱ | ۰/۳۹ |

متوسط ضریب رگرسیون پیش‌بینی ژنومی برای روش‌های GBLUP، Bayes B، RKHS، و NN به‌ترتیب ۰/۸۳، ۱/۰۸، ۱/۲۵، و ۱/۱۷ حاصل شد (جدول ۴). برای همهٔ صفات ضریب رگرسیون پیش‌بینی ژنومی حاصل برای روش‌های RKHS، NN، و بیز B مقداری بیشتر از حد بود، ولی برای روش GBLUP برای همهٔ صفات وضعیت برآورد کمتر از حد دیده شد. ضریب رگرسیون پیش‌بینی برای روش‌های BLUP برآورد بیش از حد را نشان می‌دهد و برای روش RKHS میزان برآورد کمتر از حد بود که با تحقیق حاضر متفاوت بود [۳].

در بین چهار روش آماری، روش GBLUP به نسبت سه روش دیگر دارای کمترین میانگین مربعات خطا بود، ولی به‌طور کلی میانگین مربعات خطای پیش‌بینی با روش‌های پارامتری در مقایسه با روش‌های ناپارامتری کم‌تر به‌دست آمد (جدول ۵).

متوسط عدم تعادل پیوستگی (LD) حاصل در این تحقیق (۰/۲۶۵) برای نشانگرهای همجوار (به‌ازای هر ۶۰ کیلو جفت باز یک مارکر)، در تشابه با LD جمعیت گاو هلشتاین با چیپ ۵۰ کیلوباز است [۱۲]. این سطح LD تقریباً برای رسیدن به حداکثر دقت پیش‌بینی ژنومی کفایت می‌کند [۱۱]. براساس دقت برآوردشده برای پیش‌بینی عملکرد، بیز B به نسبت دیگر روش‌ها (به‌ویژه برای صفت چربی) بهتر بود. از طرف دیگر، میانگین مربعات خطای روش GBLUP در مقایسه با روش‌های دیگر کمتر بود و روش مطلوب‌تری به نسبت دیگر روش‌هاست.

**جدول ۴. ضریب رگرسیون پیش‌بینی ژنومی برای صفات تولیدی در گاوهای هلشتاین ایران با مدل‌های متفاوت آماری**

| صفات | b(GBV,EBV) | | | |
|---|---|---|---|---|
| | GBLUP | Bayes B | RKHS | NN |
| **تولید شیر** | ۰/۸۹ | ۱/۰۷ | ۱/۴۲ | ۱/۲۳ |
| **تولید چربی** | ۰/۷۷ | ۰/۹۳ | ۱/۱۲ | ۱/۱۷ |
| **تولید پروتئین** | ۰/۷۹ | ۱/۱۱ | ۱/۲۳ | ۱/۱۱ |
| **شمارش سلول‌های سوماتیک** | ۰/۸۸ | ۱/۲۳ | ۱/۲۵ | ۱/۱۷ |

**جدول ۵. میانگین مربعات خطای پیش‌بینی ژنومی برای صفات تولیدی در گاوهای هلشتاین ایران با مدل‌های متفاوت آماری**

| صفات | MSE | | | |
|---|---|---|---|---|
| | GBLUP | Bayes B | RKHS | NN |
| تولید شیر | ۲۱۴/۱۸ | ۲۲۲/۱۵ | ۲۲۵/۲۳ | ۲۲۷/۳۴ |
| تولید چربی | ۵۶/۹۴ | ۶۲/۴۳ | ۷۰/۲۳ | ۷۲/۲۴ |
| تولید پروتئین | ۳۹/۴۳ | ۵۳/۴۰ | ۶۳/۳۴ | ۶۵/۴۵ |
| شمارش سلول‌های سوماتیک | ۰/۶۷ | ۱/۱۰ | ۱/۳۱ | ۱/۷۴ |

در روش‌هایی که تعداد نشانگرها خیلی بیشتر از تعداد حیوانات ژنوتیپ‌شده باشد، فرض این است که روش‌های انتخاب متغیر (مانند بیز B) در مقایسه با دیگر روش‌ها، شایستگی بهتری در برآورد دقت انتخاب ژنومی در داده‌های واقعی را نشان می‌دهند [۲۰] که نتایج این تحقیق در تطابق با این موضوع بود. بنابراین در مواردی که تعداد نمونهٔ دردسترس ناکافی باشد، بهتر است از روش‌های انتخاب متغیر برای برآورد تأثیرات نشانگر و درنهایت دقت برآوردها استفاده شود. در ضمن زیادبودن دقت در این تحقیق برای بعضی صفات (چربی شیر) در روش بیز B، می‌تواند نشانهٔ تأثیر ژن‌های بزرگ‌اثر را بر این صفت باشد.

کم‌بودن دقت پیش‌بینی ژنومی برای بعضی از صفات در برخی روش‌ها احتمالاً به‌دلیل ناکافی‌بودن تعداد نمونه‌ها یا به‌دلیل کاهش همپوشانی نشانگرها برای برآورد دقیق تأثیرات نشانگرها باشد. از طرف دیگر، اگرچه دقت روش بیز B به نسبت روش GBLUP تمام صفات را بیشتر برآورد کرده است، اما روش‌های بیز B و روش‌های ناپارامتری در برآورد پیش‌بینی‌ها برآوردهای کمتر از حد داشتند [۹ و ۱۰]. ولی در برخی دیگر از مطالعات روند برآورد بیش از حد در توافق با تحقیق حاضر برای صفات بودند [۱۹] و این تفاوت در نتایج، احتمالاً به‌دلیل تفاوت بین داده‌های دو تحقیق، تفاوت در روش‌های مطالعه‌شده، و اجرای این روش‌هاست.

از مهمترین پارامترهای مؤثر در برآوردهای متفاوت دقت پیش‌بینی اصلاحی ژنومی، معماری ژنتیکی صفات است [۱۳]. اگر معماری ژنتیکی صفات شامل تأثیرات متقابل اپیستازی باشد، روش‌های ناپارامتری ارزش‌های اصلاحی ژنومی را با دقت بیشتر برآورد می‌کنند، ولی اگر معماری ژنتیکی براساس تأثیرات افزایشی باشد، روش‌های پارامتری این تأثیرات را نیز برآورد می‌کنند. در تحقیق حاضر، چون روش‌های پارامتری به نسبت روش‌های ناپارامتری با دقت بیشتری ارزش‌های اصلاحی را برای صفات مطالعه‌شده برآورد کردند، می‌توان استنباط کرد که احتمالاً معماری ژنتیکی این صفات براساس تأثیرات افزایشی باشد. در توافق با این موضوع، مطالعات دیگری نیز همین روال را برای دقت پیش‌بینی ژنومی با قراردادن یا ندادن تأثیرات اپیستازی در مطالعات خود تأیید کردند [۱۳]. بیشتر نتایج از مطالعات قبلی در تأیید این امر هستند که در روش‌های پارامتری فرض این است که متغیرهای مستقل برازش‌شده در مدل از هم مجزا هستند، وقتی در مدل‌ها فقط تأثیرات افزایشی آورده می‌شود، فرض بر این است که نشانگرها، مستقل هستند که در این مورد روش‌های پارامتری دقیق‌تر از روش‌های غیرپارامتریک ارزش‌های اصلاحی ژنومی را برآورد می‌کنند [۱۳]. نتایج این تحقیق هم در مطابقت با همین فرض بود.

روش‌های پارامتری (GBLUP و Bayes B) برای بیشتر صفات در مقایسه با روش‌های ناپارامتری دقت پیش‌بینی ژنومی بیشتری را در پی داشت و در بین روش‌های پارامتری، بیز B عملکرد بهتری را نشان داد. نتیجه این که با توجه به معماری ژنتیکی صفات تولیدی در گاوهای هلشتاین ایران که احتمالاً بیشتر تحت‌تأثیر پیامدهای افزایشی باشد، می‌توان با داشتن دادهٔ ژنومی از روش‌های پارامتری برای برآورد صحت انتخاب ژنومی استفاده کرد.

## منابع


۱. عبدالهی آرپناهی ر، پاکدل ع، نجاتی جوارمی ا و مرادی شهربابک م (۱۳۹۲). مقایسهٔ روش‌های گوناگون ارزیابی ژنومیک در صفاتی با معماری ژنتیکی متفاوت. تولیدات دامی. ۱۵: ۶۵-۷۷.

2. Beckmann J and Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. TAG Theoretical and Applied Genetic. 67(1): 35-43.

3. Charfeddine N, Rodriguez-Ramilo ST, Jimenez JA, Carabano MJ and Gonzalez-Recio O (2013) Non parametric vs. GBLUP Model for genomic selection evaluation with large reference population in Holstein cattle. INTERBULL BULLETIN. 47: 23-25.

4. Daetwyler HD, Calus MPL, Pong-Wong R, De Los Campos G and Hickey JM (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics. 193: 347-365.

5. De los Campos G and Perez PR (2012) Bayesian Generalized Linear Regression BGLR. R Package. http://bglr.r-forge.rproject.org/

6. Dekkers JCM, Mathur PK and Knol EF (2010) Genetic Improvement of the Pig, pp. 390–425 in The Genetics of the Pig, Ed. 2, edited by Rothschild MF and Ruvinsky A. CABI, Cambridge, MA.

7. Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. Journal Animal Breeding Genetic. 124: 331-341.

8. De Roos APW, Schrooten C and Druet T (2011) Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. Dairy Science. 94: 4708-4714.

9. Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc JM, Barillet F, Robert-Granié C and Manfredi E ((2012) Genomic selection in the French Lacaune dairy sheep breed. Dairy Science. 95: 2723-2733.

10. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, Mason BA and Goddard ME (2012) Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. Dairy Science. 95: 4114-4129.

11. Espigolan R, Baldi F, Boligon AA, Souza FR, Gordo DG, Tonussi RL, Cardoso DF, Oliveira HN, Tonhati H, Sargolzaei M, Schenkel FS, Carvalheiro R, Ferro JA and Albuquerque LG (2013) Study of whole genome linkage disequilibrium in Nellore cattle. BMC Genomics. 14: 305.

12. Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME (2009) Invited review: genomic selection in dairy cattle: progress and challenges. Dairy Science. 92: 433-443.

13. Howard R, Carriquiry AA and Beavis WD (2014) Parametric and non parametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. G3. 4: 1027-1046.

14. Guo G, Lund MS, Zhang Y and Su G (2010) Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. Animal Breeding Genetic. 127: 423-432.

15. Gianola D and Van Kaam JBCHM (2008) Reproducing Kernel Hilbert Spaces Regression methods for genomic assisted prediction of quantitative traits. Genetics. 178: 2289-2303.

16. Goddard ME and Hayes BJ (2007) Genomic selection. Animal Breeding Genetic. 124(6): 323-330.

17. Kearsey MJ and Farquhar AGL (1998) QTL analysis in plants; where are we now? Heredity. 80: 137-142.

18. Konstantinov KV and Hayes BJ (2010) Comparison of BLUP and Reproducing kernel Hilbert spaces methods for genomic prediction of breeding values in Australian Holstein Friesian cattle. Proceedings of the 9th WCGALP, Leipzig, Germany.

19. Legarra A, Robert-Granie C, Manfredi E and Elsen JM (2008) Performance of genomic selection in mice. Genetics. 180: 611-618.

20. Meuwissen TH, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157: 1819-1829.

21. Meuwissen TH (2007) Genomic selection: marker assisted selection on a genome wide scale. Journal of Animal Breeding Genetic. 124(6): 321-322.

22. Neves H, Carvalheiro HR and O'Brien AMP, Utsunomiya YT, Do Carmo AS, Schenke FS, Sölkner J, McEwan JC, Tassell CPV, Cole JB, Da Silva MV, Queiroz SA, Sonstegard TS and Garcia JF (2014) Accuracy of genomic predictions in Bos indicus (Nellore) cattle. Genetics Selection Evolution. 46:17-25.

23. Okut H, Wu X, JM Rosa G, Bauck S, Woodward BW, Schnabel RD, Taylor JF and Gianola D (2013) Predicting expected progeny difference for marbling score in Angus cattle using artificial neural networks and Bayesian regression models. Genetics Selection Evolution. 45: 34.

24. Pérez-Rodiguez P, Gianola D, Rosa G, Weigel K and Crossa J (2013) Technical Note: An R package for fitting Bayesian regularized neural networks with applications in animal breeding. Journal of Animal Science. In press.

25. Van Raden PM, Tassell CPV, Wiggans GR, Sonstegard TS and Schnabel RD (2009) Invited review: reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science. 92: 16-24.