

مقایسه رویکرد حداکثر درست نمایی محدود شده با بیزی در برآورد اجزای واریانس ژنومی صفات پشم گوسفند مرینوس

- آذر راشدی ده‌صحرایی (نویسنده مسئول)
دانشجوی دکتری ژنتیک و اصلاح نژاد دام، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان
 - جمال فیاضی
دانشیار ژنتیک و اصلاح نژاد دام، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان.
 - رستم عبداللهی آرپناهی
استادیار ژنتیک و اصلاح نژاد دام، پردیس ابوریحان دانشگاه تهران
 - جولوس ون در ورف
استاد دانشکده کشاورزی و علوم روستائی دانشگاه نیوانگلند، آرمیدال، استرالیا
 - هدایت اله روشنفر
استاد ژنتیک و اصلاح نژاد دام، دانشگاه علوم کشاورزی و منابع طبیعی خوزستان
- تاریخ دریافت: تیر ۱۳۹۶ تاریخ پذیرش: آذر ۱۳۹۶
شماره تماس نویسنده مسئول: ۰۹۱۳۳۸۵۹۱۸۶
Email: Azar.Rashedi2010@yahoo.com

شناسه دیجیتال (DOI): 10.22092/asj.2017.114469.1480

چکیده

برآورد دقیق اجزای واریانس ژنتیکی و غیرژنتیکی با اطلاعات شجره‌ای و ژنومی، از ملزومات پیش‌بینی صحیح ارزش‌های اصلاحی می‌باشد. از آنجایی که تراشه‌های چندشکلی تک‌نوکلئوتیدی (SNP) کل ژنوم را پوشش می‌دهند، نشانگرها تمامی جایگاه‌های صفات کمی را تحت پوشش قرار داده و به طور بالقوه تمام واریانس ژنتیکی را توجیه می‌کنند. در این پژوهش از SNPهای گوسفندان مرینوس استرالیایی استفاده شد. صفات طول و قطر تار پشم مورد بررسی قرار گرفتند. برای مطالعه رابطه بین فراوانی آللی و مقدار واریانس ژنتیکی افزایشی توجیه شده، SNPها در پنج گروه مختلف از فراوانی آللی کمیاب (MAF)، طبقه بندی شدند. دو مدل آماری با آنالیز مجزای هر گروه SNP و آنالیز توأم هر پنج گروه SNP برازش شدند. آنالیزهای آماری با دو روش حداکثر درست‌نمایی محدود شده ژنومی (روش پارامتری) و بیزی (با استفاده از تکنیک نمونه‌گیری گیبس و مدل RKHS (روش نیمه‌پارامتری)) انجام شد. مقدار وراثت‌پذیری ژنومی برآورد شده توسط همه SNPها در رویکرد REML، برای قطر و طول تار پشم به ترتیب برابر ۰/۷۲ و ۰/۴۸ بود. در رویکرد بیزی این مقدار وراثت‌پذیری برای صفات مذکور به ترتیب برابر ۰/۷۴ و ۰/۴۷ برآورد گردید. در تجزیه و تحلیل مجزای گروه‌های مختلف MAF مقادیر وراثت‌پذیری ژنومی هر دو رویکرد مشابه ولی در تجزیه توأم، بین دو رویکرد تفاوت زیادی وجود داشت. بطور کلی در مدل‌های ساده نتایج دو رویکرد مشابه بوده ولی در مدل‌های پیچیده مانند آنالیز توأم نتایج دو مدل متفاوت و معنی‌دار هستند. برای بررسی این که کدام روش از اعتبار بالاتری برخوردار است نیاز به تحقیقات آتی می‌باشد.

واژه‌های کلیدی: گوسفند مرینوس، واریانس ژنومی، روش پارامتری، روش بیزی، انتخاب ژنومی

Animal Science Journal (Pajouhesh & Sazandegi) No 120 pp: 35-46

The effect of *Prosopis farcta* on the performance, some blood parameters, immune and antioxidant system of broiler chickens under heat stress conditions.

- 1- Azar Rashedi Dehsahraei. PhD Student of Genetics & Animal Breeding- Khuzestan Agriculture sciences and Natural Resources University. Mobile: +98-913-3859186. Azar.Rashedi2010@yahoo.com.
- 2- Dr. Jamal Fayazi, Associate professor (B.Sc., M.Sc., Ph.D), Department of Animal Science Khuzestan Agriculture sciences and Natural Resources University., Ahwaz, Iran
- 3- Dr. Rostam Abdollaho Arpanahi, Assistant professor, Department of Animal and Poultry Science College of Aburaihan, University of Tehran
- 4- Dr. Julius Van Der Werf. Professor, School of Environmental and Rural Science, University of New England, Armidale, Australia.
- 5- Dr. Hedayatollah Roshanfekar. Professor Department of Animal Science Khuzestan Agriculture sciences and Natural Resources University. Mollasani, Ahwaz, Iran

Received: June 2017

Accepted: November 2017

Accurate estimation of variance components using pedigree and genomic data plays a key role in prediction of breeding values. Since SNP markers in genomic selection are distributed across the genome, they may cover all quantitative traits loci and potentially explain all of genetic variation. In this study, genotype data from Merino sheep, genotyped by 50k Illumina SNP chip were used. Staple length and Fibre diameter traits were studied in this research. To study the association between allele frequency spectrum and captured additive genetic variance, all SNPs were partitioned in five MAF bins with the equal numbers of SNPs. Two statistical models including separate analysis for each category of MAF SNPs or joint analysis of all MAF groups were fitted. The analysis were performed using REML (parametric) and a Bayesian method implemented via Gibbs sampling and RKHS (**semi-parametric**) model. Using all common SNPs in REML approach, estimates of genomic heritability were 0.72 and 0.48 for Staple length and Fibre diameter, respectively. In Bayesian approach, genomic heritability for mentioned traits were 0.74 and 0.47 respectively. In the separate analysis, estimates of genomic heritability using REML and Bayesian approaches for each MAF class were similar, but in joint analysis estimates of two approaches were different. Overall, when the model is simple both approaches perform similarly while when model is complicated as joint analysis in present study, two approaches work different. Therefore, to determine which approach is more reliable, further research is required.

Key words: Merino Sheep, Genomic Variance, Parametric Method, Bayesian Method, Genomic Selection

مقدمه

(انتخاب بر اساس رکوردهای فنوتیپی و شجره‌ای) در برخی صفات مانند صفاتی که مشکلات رکوردگیری در آنها وجود دارد و یا اینکه با زنده ماندن حیوان در ارتباط هستند، باعث پیشرفت ژنتیکی بیشتری شده است. استفاده از ارزیابی ژنومی به دلیل کاهش فاصله نسلی و افزایش صحت انتخاب در بدو تولید، باعث شده است که پیشرفت ژنتیکی حدوداً به میزان ۵۰ تا ۷۰ درصد در گاو شیری افزایش یابد (schrooten و همکاران ۲۰۰۵). ویژگی مطلوب دیگر انتخاب ژنومی این است که اثرات نشانگرها بطور

تعیین ژنوتیپ نشانگرهای تک نوکلئوتیدی (SNP) و سایر نشانگرهای متراکم در ژنوم، باعث پیشرفت ژنتیکی چشمگیری در برنامه‌های اصلاح دام شده است (Meuwissen و همکاران ۲۰۰۹، Hayes ۲۰۰۹). از آنجا که نشانگرها کل ژنوم را پوشش می‌دهند، تمامی جایگاه‌های صفات کمی تحت پوشش نشانگرها قرار گرفته و این روش به طور بالقوه توانایی توجیه تمام واریانس ژنتیکی را دارا می‌باشد (تیموریان و همکاران ۱۳۹۴). انتخاب ژنومی در مقایسه با روش های سنتی

۷۵ درصد از واریانس ژنومی به وسیله نشانگرهایی با فراوانی کمتر از ۰/۳ شرح داده شد. Pimentel و همکاران (۲۰۱۱) برای صفات تولید و ترکیب شیر در گاوهای شیری، Jensen و همکاران (۲۰۱۲) برای صفات مرتبط با تولید و تناسب اندام در گاوهای شیری، Lee و همکاران (۲۰۱۳) برای بیماری آلزایمر، Ogawa و همکاران (۲۰۱۶) برای وزن بدن گاو سیاه ژاپنی و Yang و همکاران (۲۰۱۰) برای قد انسان با استفاده از نشانگرهای SNP مطالعاتی را برای برآورد مقدار تنوع ژنومی انجام دادند. هدف از انجام پژوهش حاضر، برآورد اجزای واریانس ژنومی با استفاده از نشانگرهای SNP و طبقه‌بندی نشانگرها بر اساس فراوانی آللی کمیاب (MAF)، برای صفات طول تار پشم^۱ و قطر تار پشم^۲ در گوسفند مریوس استرالیا می‌باشد. مقایسه دو رویکرد آماری مختلف پارامتری شامل REML و بیزی از دیگر اهداف این پژوهش هستند.

مواد و روش‌ها

برای انجام این پژوهش از رکوردهای فنوتیپی گوسفندان مریوس استرالیا (۱۲۲۹ نر و ۹۶۰ ماده)، استفاده شد. دام‌ها متعلق به ۱۱ خانواده ناتنی پدری بوده که همگی در ایستگاه مزرعه فالکینر استرالیا (FMFS)^۴ متولد شدند. صفات قطر (۹۵۴ رکورد) و طول تار پشم (۹۵۴ رکورد) در این پژوهش مورد بررسی قرار گرفتند. آمار توصیفی صفات مورد مطالعه در جدول ۱ آورده شده است.

همزمان برآورد می‌شود (Hayes, ۲۰۰۷) و همچنین برای هر قطعه کروموزومی می‌توان ارزیابی متفاوت با دیگری قائل شد و از این اطلاعات در انتخاب حیوانات استفاده نمود. ارزیابی ژنومی با تکیه بر اصول فوق و با اعتقاد به اینکه تمام واریانس ژنتیکی صفت توسط نشانگرهایی که با QTLها و ژن‌های مؤثر بر صفات در عدم تعادل پیوستگی هستند، توجیه می‌شود قابل تفسیر است (قائم‌ی و همکاران، ۱۳۹۱).

تخمین مؤلفه‌های (کو) واریانس استفاده وسیعی در اصلاح نژاد دام دارد و این مؤلفه‌ها باید به طور دقیق برآورد گردند (دیمی غیاث‌آبادی، ۱۳۹۱). در این ارتباط از قویترین روش‌های برآورد اجزای واریانس، در اصلاح نژاد دام می‌توان روش حداکثر درست‌نمایی محدود شده (REML^۱) و روش بیزی مبتنی بر تکنیک نمونه‌گیری گیبس را نام برد. در روش REML پیش فرض اساسی این است که نمونه‌ها از یک جامعه با توزیع نرمال هستند. این روش برای اثر انتخاب در جامعه تصحیح انجام می‌دهد و مؤلفه‌های واریانس را از طریق تکرار و همگرایی به دست می‌آورد (دیمی غیاث‌آبادی، ۱۳۹۱). نمونه‌گیری گیبس یک روش همانندسازی عددی است که اجازه می‌دهد از توزیع‌های مشترک یا حاشیه‌ای حتی در حالتی که چگالی‌های مناسب نتواند به درستی شکل بگیرد، استدلال‌هایی ایجاد شود (Geman and Geman, ۱۹۸۴).

در سال‌های اخیر مطالعات زیادی با استفاده از تراشه‌های SNP، برای شناسایی تنوع ژنتیکی، انجام گرفته است. در پژوهشی روی گاو سیاه ژاپنی، توسط Uemoto و همکاران (۲۰۱۵)، مقدار وراثت‌پذیری QTL در جمعیت‌هایی با اندازه متفاوت، تقریباً مشابه (۰/۴) برآورد شد. اما در پژوهش این محققین مقدار خطای استاندارد وراثت‌پذیری با کاهش اندازه جمعیت، افزایش یافت. Abdollahi-Arpanahi و همکاران (۲۰۱۴)، اثر فراوانی آللی را روی پیش‌بینی صفات کمی در جوجه‌های گوشتی بررسی نمودند. این محققین گزارش کردند که ۷۵ درصد از واریانس ژنومی برای صفات وزن بدن و عضله سینه توسط نشانگرهایی با فراوانی کمتر از ۰/۲ توجیه می‌شود. برای صفت تولید تخم‌مرغ نیز

¹ Restricted Estimation Maximum Likelihood

² Staple length

³ Fibre diameter

⁴ Falkiner Memorial Field Station

جدول ۱- آمار توصیفی صفات مورد بررسی

بیشین	کمینه	انحراف معیار	میانگین	تعداد	صفت (واحد)
۳۰/۴	۱۴/۲	۲/۱۸	۱۹/۱۹	۹۵۴	قطر تار پشم (میکرون)
۹۵	۵۱	۱۰/۹۱	۸۱/۰۲	۹۵۴	طول تار پشم (میلیمتر)

جدول ۲- خصوصیات تراشه SNP مورد استفاده

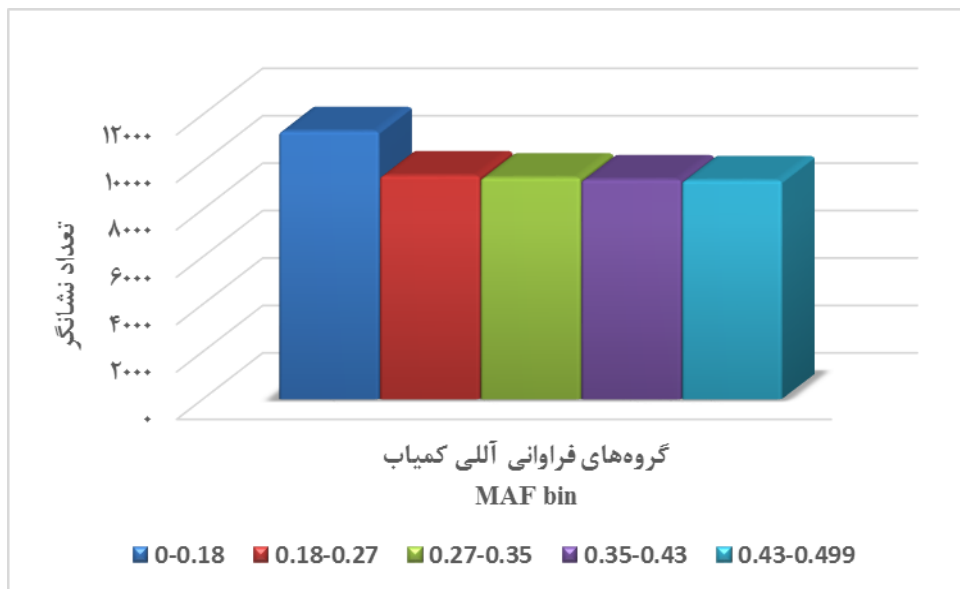
۴۸۵۹۹	تعداد کل SNPها
.	SNPهای با موقعیت ناشناخته
۳۸۴	SNPهای با فراوانی آللی کمیاب (>0.01)
۸۷۳	SNPهایی در انحراف از تعادل هاردی-واینبرگ ($p < 10^{-6}$)
.	نرخ نشانگر گمشده (>0.05)
۴۷۳۴۲	SNPهای مورد استفاده
.	افراد با نرخ ژنوتیپ گمشده (>0.05)
۵۵/۱۸۸	میانگین فاصله SNPها (کیلو جفت باز)
۱۱۸۰/۸۸۶	بیشترین فاصله (کیلو جفت باز)
۴/۵۴۸	کمترین فاصله (کیلو جفت باز)

مختلط خطی برآورد شد. برای مطالعه رابطه بین فراوانی آللی و مقدار واریانس ژنتیکی افزایشی، SNPها در پنج گروه مختلف از فراوانی آللی کمیاب (MAF)، با تعداد تقریباً برابر در هر گروه، طبقه بندی شدند. دسته بندی پنج گروه مختلف نشانگرها با توجه به فراوانی آللی کمیاب بدین صورت بود (۰/۱۸-۰/۲۷، ۰-۰/۱۸، ۰/۳۶-۰/۴۳، ۰/۲۷-۰/۳۶، ۰/۴۳-۰/۴۳ و ۰/۴۳-۰/۴۹۹). تعداد نشانگرها در گروه‌های مختلف MAF، در شکل ۱ نشان داده شده است.

از مجموعه اولیه SNPها (۵۰۰۰۰ SNP)، ۴۸۵۹۹ تای آنها نشانگر اتوزومی بودند. پس از کنترل کیفیت، ۳۸۴ نشانگر به خاطر فراوانی آللی کمیاب ($MAF^5 < 0.01$) و ۸۷۳ نشانگر به دلیل انحراف از تعادل هاردی-واینبرگ ($p \text{ value} < 10^{-6}$) حذف شدند و ۴۷۳۴۲ نشانگر در آنالیزها باقی ماندند. توزیع SNPها در طول ۲۶ جفت کروموزوم همولوگ، با فاصله ۵۵/۱۸۸ کیلو جفت باز به طور متوسط بین SNPهای مجاور (۴/۵۴۸ کیلو جفت باز بین نزدیک‌ترین SNP و ۱۱۸۰/۸۸۶ کیلو جفت باز بین دورترین SNP) بود. اگر چه بیشتر SNPها بر روی کروموزوم‌های طویل‌تر بودند، اما متوسط فاصله بین دو SNP در هر کروموزوم تقریباً مشابه بود. ویرایش و کنترل کیفیت داده‌ها با استفاده از نرم‌افزار PLINK در سیستم عامل لینوکس (Purcell و همکاران، ۲۰۰۷) انجام گرفت. خلاصه‌ای از تجزیه و تحلیل نشانگرها در جدول ۲ نشان داده شده است.

بعد از ویرایش و کنترل کیفیت داده‌ها، مقدار تنوع ژنتیکی توجه شده توسط همه SNPها به طور همزمان، با استفاده از یک مدل

⁵ Minor Allele Frequency



شکل ۱- تعداد نشانگرهای SNP در گروه‌های مختلف فراوانی آللی کمیاب (MAF)

مختلف استفاده شد. رویکرد اول: روش حداکثر درست‌نمایی محدود شده ($REML^6$) به وسیله نرم‌افزار آنالیز پویس ژنومی صفات پیچیده ($GCTA^7$)، انجام شد. GCTA در واقع برای برآورد نسبتی از واریانس فنوتیپی توجیه شده توسط SNP‌های ژنوم یا کروموزوم گسترده برای صفات پیچیده، طراحی شده است (روش حداکثر درست‌نمایی محدود شده ژنومیک ($GREML^8$))، (Yang و همکاران، ۲۰۱۰). رویکرد دوم: از روش بیزی برای برآورد اجزای واریانس ژنومی با استفاده از تکنیک نمونه‌گیری گیبس و مدل آماری RKHS (Gianola and de los Campos، ۲۰۱۳) از طریق بسته BGLR⁹ در محیط R، استفاده شد (Pérez and de los Campos، ۲۰۱۳). در پژوهش حاضر، طول زنجیره نمونه‌گیری گیبس ۱۰۰۰۰۰ تکرار، تعداد تکرارهای قلق‌گیری ۱۰۰۰۰ و فاصله نمونه‌گیری ۵ در نظر گرفته شد. در تجزیه و تحلیل داده‌ها با استفاده از رویکرد REML مدل آماری در آنالیز مجزا به صورت زیر بود:

$$y = 1\mu + g + e \quad (2)$$

به منظور برآورد نسبتی از تنوع ژنتیکی افزایشی در هر یک از پنج گروه MAF، تجزیه و تحلیل به دو روش مختلف انجام گرفت. در روش اول (آنالیز جداگانه) پنج مدل با اثر تصادفی حیوان و ساختار واریانس-کواریانس ماتریس روابط خویشاوندی هر یک از پنج گروه مختلف فراوانی آللی کمیاب (MAF) برآزش داده شد. در روش دوم (آنالیز توأم) یک مدل که پنج گروه مختلف MAF، به عنوان اثر ژنتیکی تصادفی مختلف هر کدام با یکی از ماتریس‌های رابطه خویشاوندی فوق‌الذکر در ارتباط بود برآزش داده شد. از این مدل، پنج جزء واریانس ژنتیکی افزایشی متمایز که هر کدام به یکی از گروه‌های فراوانی آللی کمیاب اختصاص داشت، به طور همزمان محاسبه گردید.

آنالیزهای آماری: در ابتدا فنوتیپ‌ها برای اثرات ثابت طبق مدل زیر تصحیح شدند و از فنوتیپ تصحیح شده به عنوان متغیر پاسخ استفاده شد.

$$y_{ijkl} = \mu + S_i + T_j + M_k + e_{ijkl} \quad (1)$$

در این مدل y_{ijkl} : هر یک از مشاهدات برای صفت مورد مطالعه، μ : میانگین جامعه، S_i : اثر آمین جنس بره، T_j : اثر آمین تیپ تولد بره، M_k : اثر آمین ماه تولد بره و e_{ijkl} : اثرات باقیمانده.

در این مطالعه برای برآورد اجزای واریانس ژنومی از دو رویکرد

⁶ Restricted Estimation Maximum Likelihood

⁷ Genome-wide Complex Trait Analysis

⁸ Genomic Restricted Estimation Maximum Likelihood

⁹ Bayesian Generalized Linear Regression

صورت زیر حداکثر می کند:

$$l(\alpha|\lambda) = (y - K\alpha)'(y - K\alpha) + \lambda\alpha'K\alpha \quad (6)$$

در رابطه بالا $\alpha \sim N(0, K^{-1}\sigma_g^2)$ ، واریانس ژنومی افزایشی σ_g^2 ، واریانس ژنومی افزایشی توجیه شده به وسیله نشانگرها و λ : یک فاکتور تنظیم کننده است. مقادیر ارزش های ژنتیکی افزایشی از رابطه $g = K\alpha$ بدست می آید. یکی از مزایای استفاده از RKHS این است که می تواند ابعاد معادلات را از $m \times m$ یعنی صدها هزار SNP به تعداد $n \times n$ یعنی برابر با تعداد مشاهدات کاهش دهد. بنابراین می توان معادله را به شکل ماتریسی زیر نوشت:

$$y = 1\mu + G\alpha + e \quad (7)$$

تمامی اجزای این معادله قبلا در معادله (۲) تعریف شده است.

تجزیه و تحلیل توأم با پنج کرنل افزایشی ($G_1 + G_2 + \dots + G_5$) در آنالیز توأم پنج گروه مختلف MAF از مدل آماری زیر استفاده شد:

$$y = 1\mu + \sum_{t=1}^h K_t \alpha_t + e \quad (8)$$

تمامی اجزای معادله (۷) قبلا در معادله (۴) و (۵) تعریف شده است به استثنای $\alpha_t \sim (0, K_t^{-1}\sigma_{gt}^2)$ و $t=1, 2, \dots, 5$ یک بردار رگرسیون در RKHS برای گروه های MAF α_t ، K_t : ماتریس خویشاوندی افزایشی ژنومی برای t امین گروه MAF، σ_{gt}^2 : واریانس توجیه شده توسط SNP ها در t امین گروه MAF است. بنابراین تابع چگالی توأم پنج بردار تصادفی و باقیمانده به صورت زیر می باشد:

$$p(e, g_1, \dots, g_t | \sigma_e^2, \sigma_{g_1}^2, \dots, \sigma_{g_t}^2) = N(e | I\sigma_e^2) \prod_{t=1}^h N(g_t | 0, K_t \sigma_{g_t}^2) \quad (9)$$

و ماتریس واریانس-کواریانس برابر:

$$V = \sum_{t=1}^h K_t \sigma_{g_t}^2 + I\sigma_e^2 \quad (10)$$

است. در اینجا فرض بر این است که هیچ کواریانس بین اثرات SNP ها در گروه های مختلف وجود ندارد.

در این مدل: y : بردار مشاهدات تصحیح شده برای اثرات ثابت، μ : میانگین، 1 : بردار واحد، $g \sim N(0, G\sigma_g^2)$: بردار تصادفی ارزش ژنتیکی افزایشی، G : ماتریس روابط خویشاوندی ژنومی افزایشی، σ_g^2 : واریانس ژنومی توجیه شده توسط m نشانگر، $e \sim N(0, I\sigma_e^2)$: بردار باقیمانده های مدل و σ_e^2 : واریانس باقیمانده است، I : ماتریس واحد با ابعادی $n \times n$ و n برابر با تعداد حیوان دارای رکورد است. در این معادله عناصر ماتریس G با استفاده از رابطه زیر محاسبه شدند (Van Raden, ۲۰۰۸).

$$G_{jj} = \frac{1}{m} \sum_{i=1}^m \frac{(x_{ij}-2p_i)(x_{ij'}-2p_i)}{2p_i(1-p_i)} \quad (3)$$

در این معادله x_{ij} : تعداد کپی آلل کمیاب برای i امین نشانگر SNP در j امین فرد، p_i : فراوانی آلل کمیاب، m : تعداد نشانگرهای SNP می باشد. در مطالعات به این مدل بهترین پیش بینی نارایب خطی ژنومی ژنومی^{۱۰} (GBLUP) می گویند. در تجزیه و تحلیل توأم، واریانس ژنتیکی توجیه شده توسط SNP ها، با استفاده از مدل خطی زیر محاسبه گردید:

$$y = 1\mu + \sum_{t=1}^h g_t + e \quad (4)$$

در این معادله $g_t \sim (0, G_t \sigma_{gt}^2)$ ، $t = 1, 2, \dots, h$. ساختار واریانس-کواریانس اثرات تصادفی به صورت زیر است.

$$V = \sum_{t=1}^h G_t \sigma_{g_t}^2 + I\sigma_e^2 \quad (5)$$

در این معادله h : تعداد گروه های MAF (در پژوهش حاضر $h=5$). فرض بر این است که هیچ کواریانسی بین گروه های مختلف MAF وجود ندارد.

در رویکرد بیزی مدل های آماری مشابه با رویکرد REML بودند به استثناء اینکه اثرات تصادفی ژنتیکی افزایشی از ضرب ماتریس خویشاوندی کرنل خطی (K) در α حاصل شدند. در اینجا K یک ماتریس کرنل $n \times n$ بوده که توسط کواریت های SNP نمایه شده است و $K \sim XX^T$ بوده که در آن X : ماتریس ژنوتیپ SNP ها است. K از ماتریس X مرکزی شده و استاندارد بدست می آید، سپس طبق پیشنهاد Van Raden (۲۰۰۸) و Yang و همکاران (۲۰۱۰) بر تعداد SNP تقسیم شده است. α بردار ضرایب رگرسیون RKHS است که تابع درست نمایی را به

¹⁰ Genomic Best Linear Unbiased Prediction (GBLUP)

نتایج و بحث

در یک گله مریونس، مقدار وراثت پذیری را برای صفات طول و قطر تار پشم به ترتیب ۰/۳۷ و ۰/۶۸ برآورد نمودند. تفاوت مشاهده شده در اجزای واریانس به دست آمده در روش ژنومی و شجره‌ای می‌تواند به دلیل تفاوت ساختار داده‌ها و مدل‌های آماری استفاده شده مبتنی بر شجره باشد (Lee و همکاران، ۲۰۱۳). Safari و همکاران (۲۰۰۷) مقدار وراثت پذیری مستقیم را برای قطر تار پشم در گوسفند مریونس استرالیا، با استفاده از اطلاعات شجره‌ای ۰/۶۸ برآورد نمودند. محققین در مطالعه‌ای بیان داشتند که مقایسه واریانس افزایشی توجیه شده به وسیله همه SNPها و کل تنوع ژنتیکی برآورد شده توسط مطالعات مبتنی بر شجره، ساده نیست، با این حال برآوردهای حاصل از مطالعات شجره‌ای، به دلیل وجود تنوع ژنتیکی غیرافزایشی و تأثیر عوامل محیطی، می‌تواند اریب باشد (Lee و همکاران، ۲۰۱۳).

اجزای واریانس ژنومی برای صفات مورد بررسی با استفاده از کل نشانگرها، در دو روش حداکثر درست نمایی محدود شده ژنومیک و بیزی در جدول ۳ نشان داده شده است. مقدار وراثت پذیری ژنومی برآورد شده توسط همه SNPها در رویکرد حداکثر درست نمایی محدود شده، برای قطر و طول تار پشم به ترتیب برابر ۰/۷۲ و ۰/۴۸ بود. در رویکرد بیزی این مقدار وراثت پذیری برای صفات مذکور به ترتیب برابر ۰/۷۴ و ۰/۴۷ برآورد گردید. علت تشابه برآوردهای دو رویکرد آماری ساده بودن مدل آماری و همچنین کافی بودن اندازه داده‌ها می‌باشد. با این وجود به دلیل خطای استاندارد پایین تر برآوردهای حاصل از رویکرد بیزی به نظر میرسد نتایج حاصل از رویکرد بیزی دقیق تر و معتبرتر از رویکردهای دیگر باشند (راشیدی ده صحرائی و همکاران، ۱۳۹۲). Matebesi و همکاران (۲۰۰۹) با استفاده از اطلاعات شجره‌ای

جدول ۳- اجزای واریانس ژنومی و پارامترهای ژنتیکی با استفاده از آنالیز تک متغیره و دو رویکرد حداکثر درست نمایی محدود شده و بیزی با استفاده از تمام نشانگرها

صفت (واحد)	روش	$\sigma_g^2 \pm SE$	$\sigma_p^2 \pm SE$	$\sigma_e^2 \pm SE$	$h_m^2 \pm SE$
قطر تار پشم (میکرون)	روش REML	۲/۴۸±۰/۳۵	۳/۴۷±۰/۱۹	۰/۹۹±۰/۲۲	۰/۷۲±۰/۰۷
	روش بیزی	۲/۵۶±۰/۳۴	۳/۴۵±۰/۱۹	۰/۸۹±۰/۲۱	۰/۷۴±۰/۰۳
طول تار پشم (میلیمتر)	روش REML	۵۶/۵۳±۱۲/۶۵	۱۱۷/۷۱±۶/۵۷	۶۱/۱۸±۹/۰۶	۰/۴۸±۰/۰۹
	روش بیزی	۵۴/۹۴±۱۳/۰۵	۱۱۵/۷۵±۶/۶۳	۶۰/۸۱±۹/۵۷	۰/۴۷±۰/۰۳

σ_g^2 : واریانس افزایشی ژنومی، σ_p^2 : واریانس فنوتیپی، σ_e^2 : واریانس باقیمانده، h_m^2 : وراثت پذیری ژنومی، SE: مقدار خطای معیار

دست آمده از پنج گروه مختلف MAF، در آنالیزهای جداگانه و توأم با یکدیگر متفاوت بود. در آنالیز جداگانه، واریانس‌های مربوط به هر گروه، به دلیل عدم تعادل پیوستگی ژن‌ها (LD) اریب بود. برای هر دو صفت، در هر دو روش GREML و RKHS، مقادیر برآورد شده در آنالیزهای جداگانه، برای همه گروه‌ها، بیشتر از مقادیر به دست آمده در آنالیز توأم بود. در

جهت در نظر گرفتن سهم نشانگرها در طیف آللی مختلف در توجیه واریانس ژنتیکی برای در صفات مورد مطالعه پنج گروه مختلف MAF تعریف شد. مقدار وراثت پذیری ژنومی مرتبط با گروه‌های مختلف MAF، برای طول و قطر تار پشم در دو آنالیز جداگانه و توأم، با استفاده از دو روش مختلف مورد استفاده (بیزی و REML) در شکل ۱ نشان داده شده است. برآوردهای به

تجزیه و تحلیل جداگانه گروه‌های مختلف MAF، در هر دو روش، مقدار واریانس ژنومی به دست آمده برای گروه‌های مختلف تا حدی مشابه بود. در آنالیز توأم، بین واریانس ژنتیکی شناسایی شده توسط زیرگروه‌های مختلف، تنوع بالایی وجود داشت. مجموع وراثت‌پذیری برآورد شده در آنالیز توأم، مشابه مقدار به دست آمده از کل SNPها، برای هر دو صفت و در هر دو رویکرد بود، اما در آنالیز جداگانه مجموع وراثت‌پذیری در پنج گروه مختلف فراوانی آللی کمیاب، بسیار بیشتر از مقدار حاصل از کل SNPها بود. بنابراین در ادامه متن مقاله، تنها نتایج حاصل از تجزیه و تحلیل توأم شرح داده می‌شود. اجزای واریانس ژنومی و پارامترهای ژنتیکی مرتبط با گروه‌های مختلف MAF، برای صفات مورد مطالعه با استفاده از روش REML و بیزی به ترتیب در جداول ۴ و ۵ نشان داده شده‌اند. مقدار وراثت‌پذیری ژنومی قطر تار پشم در پنج گروه MAF با رویکرد REML در دامنه ۰ تا حدود ۰/۴ متغیر بود. بیشترین مقدار برآورد شده در گروه پنجم با فراوانی آللی کمیاب ۰/۴۳-۰/۵ بود و کمترین مقدار در گروه اول با فراوانی آللی کمیاب ۰-۰/۱۸ بود (شکل ۱ و جدول ۴). در آنالیز بیزی نیز بین پنج گروه مختلف، در برآورد وراثت‌پذیری صفت قطر تار پشم تفاوت وجود داشت. بیشترین مقدار وراثت‌پذیری در گروه پنجم با فراوانی آللی ۰/۴۳-۰/۵ مشاهده شد (حدود ۰/۲). کمترین مقدار وراثت‌پذیری در گروه سوم بود که حدود ۰/۰۹۹ برآورد گردید (شکل ۱ و جدول ۵). برای صفت طول تار پشم، الگوی توزیع واریانس ژنتیکی توجه شده به وسیله SNPها، بین پنج گروه MAF، دارای نوسان بوده و از ۰ در گروه‌های دوم و چهارم تا ۰/۳۳ در گروه اول متغیر بود (رویکرد REML، جدول ۴ و شکل ۱). در رویکرد بیزی برای طول تار پشم مقدار وراثت‌پذیری از صفر در گروه‌های دوم و چهارم تا حدود ۰/۱۵ در گروه سوم متغیر بود (جدول ۵ و شکل ۱). با توجه به این‌که در جمعیت‌های دامی انتخاب صورت می‌گیرد توزیع واریانس ژنتیکی بر حسب فراوانی آللی متغیر می‌باشد. بین دو روش آماری استفاده شده در پژوهش حاضر، از آنجایی که در رویکرد بیزی، در گروه‌های مختلف MAF، یک

توزیع پیشین مشترک به واریانس‌ها اختصاص داده می‌شود و انعطاف‌پذیری بیشتری دارد، نتایج حاصل از این رویکرد قطعیت بیشتری داشته و از صحت بالایی برخوردارند. دیمی غیاث‌آبادی و همکاران (۱۳۹۱) در مطالعه‌ای روی مرغ‌های بومی استان فارس، دو روش آماری بیزی و REML را مورد مقایسه قرار داده و عنوان کردند که روش بیزی بسیار دقیق است و پارامترهای ژنتیکی را با دقت بالایی پیش‌بینی می‌کند. این محققین روش آنالیز بیزی را قدرتمندتر از روش REML معرفی نمودند. در پژوهش حاضر در روش بیزی حدود ۸۵ درصد از واریانس ژنتیکی قطر تار پشم و ۷۴ درصد از واریانس ژنتیکی طول تار پشم توسط SNPهای رایج با $MAF > 0.18$ ، توجیه شدند و مقدار اندکی از واریانس، مربوط به SNPهای غیر رایج و واریانت‌های کمیاب بود. عدم تعادل لینکاژ (LD)، بین واریانت‌های علی^{۱۱} با فراوانی پایین و SNPهای ژنوتیپ شده رایج، پایین است (Wray, ۲۰۰۵ و Lee و همکاران، ۲۰۱۲)، بنابراین مشاهده شده که SNPهای رایج با $MAF > 0.2$ ، بخش قابل توجهی از تنوع را در نفاطی که واریانت‌های علی رایج، دارای LD بالایی با این SNPها هستند، توجیه می‌کند (Lee و همکاران، ۲۰۱۲). در یک تحقیق در گاو سیاه ژاپنی و روی صفت وزن بدن، مقدار نسبتاً بالایی از واریانس ژنتیکی افزایشی، هنگام استفاده از SNPهای با فراوانی آللی کمیاب (MAF) بیشتر از ۰/۲ توجیه شد (Ogawa و همکاران، ۲۰۱۶) که با نتایج این مطالعه مطابقت داشت. Abdollahi-Arpanahi و همکاران (۲۰۱۴) در مطالعه‌ای روی جوجه‌های گوشتی گزارش نمودند که ۷۵ درصد از واریانس ژنومی برای صفات وزن بدن و عضله سینه توسط نشانگرهای با فراوانی کمتر از ۰/۲ توجیه می‌شود که با نتایج به دست آمده در این پژوهش مغایرت داشت. در مطالعات دیگر روی بیماری آلزایمر، اندومترروز و اسکروز ۹۰ درصد از واریانس ژنتیکی توسط SNPهای رایج با $MAF > 0.1$ (Lee و همکاران، ۲۰۱۳) و در استعداد ابتلا به بیماری‌های سرطان، دیابت نوع یک و دیابت نوع دو، بیشترین توجیه واریانس ژنتیکی توسط SNPهای با $MAF > 0.2$ بوده است (Park و همکاران، ۲۰۱۱). تمرکز روی طیف فراوانی آللی

¹¹causal variant

نتیجه گیری

نتایج نشان داد، تقریباً همه واریانس ژنتیکی توسط نشانگرهای SNP توجیه می‌شود. سهم گروه‌های مختلف SNP‌های با فراوانی آلی کمیاب در توجیه واریانس ژنتیکی برای دو صفت مورد بررسی متفاوت بوده و به طور کلی بخش قابل توجهی از واریانس ژنتیکی، توسط SNP‌های با $MAF > 0.20$ توجیه شد. زمانی که آنالیزهای جداگانه انجام شد، اجزای واریانس به دست آمده در هر دو روش REML و بیزی یکسان بودند ولی در آنالیز توأم و استفاده از مدل پیچیده‌تر، اجزای واریانس محاسبه شده در دو روش متفاوت بود. برای تعیین اعتبار دو روش REML و بیزین به مطالعات بیشتری نیاز است.

قدردانی

بدینوسیله از دانشگاه کشاورزی و منابع طبیعی رامین خوزستان بخاطر حمایت‌های مالی، از پردیس ابوریحان دانشگاه تهران به دلیل همکاری صمیمانه و از دانشگاه نیوانگلند استرالیا به دلیل اختیار گذاشتن اطلاعات داده‌ها تشکر می‌گردد.

واریانس‌های علی، برای یک صفت خاص، اثراتی را که این واریانس‌ها ممکن است بر صفات دیگر داشته باشند نادیده می‌گیرد (Lee و همکاران ۲۰۱۳ و Abdollahi-Arpanahi و همکاران ۲۰۱۴). بنابراین، حجم نمونه بسیار بزرگ و پوشش بهتر واریانس‌های با فراوانی پایین لازم است تا بتوان استنتاج قوی‌تر و قابل اعتمادتری به دست آورد (Lee و همکاران، ۲۰۱۳). بنابراین نتایج به دست آمده در مطالعه حاضر می‌تواند با پلیوتروپی (یک ژن دو یا چند اثر فنوتیپی مشخص و متفاوت ایجاد می‌کند) صفات پیچیده متعدد، که تحت یک سبک معماری پلی ژنیک هستند، سازگار باشد. با توجه به نتایج به دست آمده، در مدل‌های ساده نتایج دو رویکرد مشابه بوده ولی در مدل‌های پیچیده مانند آنالیز توأم در این مطالعه، نتایج دو مدل متفاوت و معنی‌دار بودند. برای بررسی این که کدام روش از اعتبار بالاتری برخوردار است، به نظر می‌رسد نیاز به تحقیقات آتی می‌باشد تا بتوان با اطمینان بیشتری در این مورد نتیجه‌گیری کرد.

جدول ۴- اجزای واریانس ژنومی و پارامترهای ژنتیکی با استفاده از آنالیز توأم مرتبط با گروه‌های مختلف فراوانی آلی کمیاب در رویکرد REML

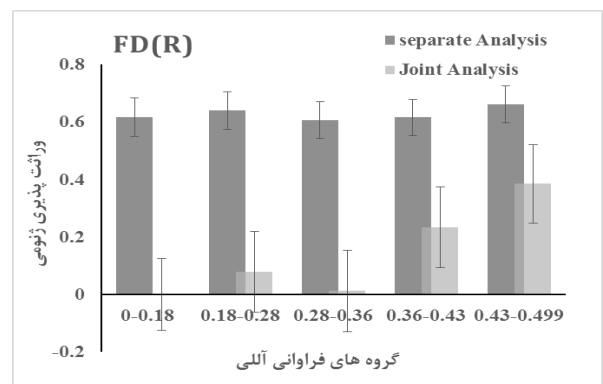
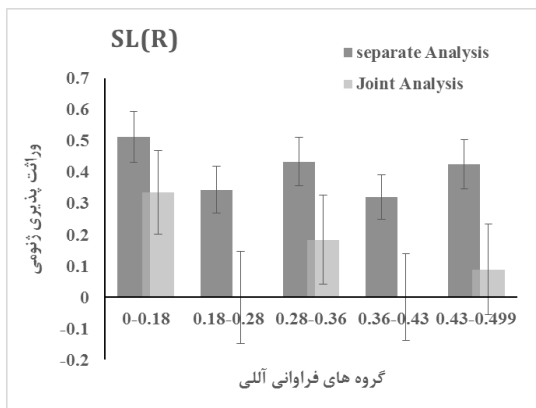
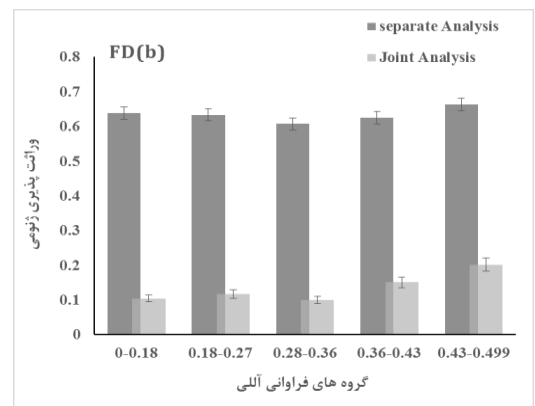
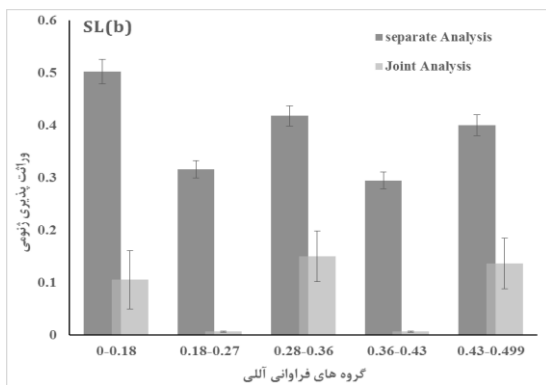
صفت	MAF-bin	σ_g^2	σ_p^2	σ_e^2	h_m^2
قطر تار پشم (میکرون)	۰-۰/۱۸	۰/۰۰۰۰۱	۳/۴۸۹ ^۱	۱/۰۱۹	۰۰۰
	۰/۱۸-۰/۲۸	۰/۲۷۳			۰/۰۷۸
	۰/۲۸-۰/۳۶	۰/۰۴۳			۰/۰۱۲
	۰/۳۶-۰/۴۳	۰/۸۱۰			۰/۲۳۲
	۰/۴۳-۰/۴۹۹	۱/۳۴۳			۰/۳۸۵
طول تار پشم (میلیمتر)	۰-۰/۱۸	۳۸/۲۰۶	۱۱۳/۹۲۰	۴۴/۵۹۲	۰/۳۳۵
	۰/۱۸-۰/۲۸	۰/۰۰۰۰۱			۰۰۰
	۰/۲۸-۰/۳۶	۲۰/۹۷۸			۰/۱۸۴
	۰/۳۶-۰/۴۳	۰/۰۰۰۰۱			۰۰۰
	۰/۴۳-۰/۴۹۹	۱۰/۱۴۳			۰/۰۸۹

MAF-bin: گروه‌های مختلف فراوانی آلی کمیاب، σ_g^2 : واریانس افزایشی ژنومی، σ_p^2 : واریانس فنوتیپی، σ_e^2 : واریانس باقیمانده، h_m^2 : وراثت‌پذیری ژنومی.
^۱ به دلیل اینکه آنالیز به صورت توأم بوده و پنج گروه MAF به صورت همزمان در مدل برازش می‌شوند، فقط یک مقدار واریانس باقی مانده و فنوتیپی برآورد می‌شود.

جدول ۵- اجزای واریانس ژنومی و پارامترهای ژنتیکی با استفاده از آنالیز توأم مرتبط با گروه‌های مختلف فراوانی آللی کمیاب در رویکرد بیزی

h_m^2	σ_e^2	σ_p^2	σ_g^2	MAF-bin	صفت
۰/۱۰۴	۱/۱۲۱	۳/۴۳۲ ^۱	۰/۳۵۶	۰-۰/۱۸	قطر تار پشم (میکرون)
۰/۱۱۶			۰/۴۰۰	۰/۱۸-۰/۲۸	
۰/۰۹۹			۰/۳۴۳	۰/۲۸-۰/۳۶	
۰/۱۵۱			۰/۵۱۷	۰/۳۶-۰/۴۳	
۰/۲۰۱			۰/۶۹۲	۰/۴۳-۰/۴۹۹	
۰/۱۰۵	۶۷/۹۰۳	۱۱۳/۶۸۳	۱۱/۹۵۶	۰-۰/۱۸	طول تار پشم (میلیمتر)
۰/۰۰۶			۰/۶۵۳	۰/۱۸-۰/۲۸	
۰/۱۵۰			۱۷/۰۹۱	۰/۲۸-۰/۳۶	
۰/۰۰۶			۰/۶۸۳	۰/۳۶-۰/۴۳	
۰/۱۳۷			۱۵/۵۷۷	۰/۴۳-۰/۴۹۹	

MAF-bin: گروه‌های مختلف فراوانی آللی کمیاب، σ_g^2 : واریانس افزایشی ژنومی، σ_p^2 : واریانس فنوتیپی، σ_e^2 : واریانس باقیمانده، h_m^2 : وراثت‌پذیری ژنومی
^۱ به دلیل اینکه آنالیز به صورت توأم بوده و پنج گروه MAF به صورت همزمان در مدل برازش می‌شوند، فقط یک مقدار واریانس باقی مانده و فنوتیپی برآورد می‌شود.



شکل ۱- وراثت‌پذیری ژنومی برآورد شده با استفاده از SNPها در پنج گروه مختلف MAF. FD(b) برای قطر تار پشم با روش بیزی، SL(b) برای طول تار پشم با روش بیزی، FD(R) برای قطر تار پشم با روش REML، SL(R) برای طول تار پشم با روش REML مبتنی بر دو روش آنالیز جداگانه و توأم

منابع

- Iowa State university.
- Hayes, B.J., P. M. Visscher, and M.E. Goddard. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetic Research*. 91:47-60.
- Jensen, J., Su, G. and Madsen, P. (2012). Partitioning additive genetic variance into genomic and remaining polygenic components for complex traits in dairy cattle. *BMC Genetics*. 13, 44.
- Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E. and et al. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genetics*. 44: 247-250.
- Lee, S.H., Harold, D., Nyholt, D.R., Goddard, M.E., Zondervan, K.T., Williams, J. et al. (2013). Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Human Molecular Genetics*. 22: 832-841.
- Matebesi, P.A., Cloete, S.W.P. and Van Wyk, J.B. (2009). Genetic parameter estimation of 16-month live weight and objectively measured wool traits in the Tygerhoek Merino flock. *South African Journal of Animal Science*. 2009, 39 (1): 73-82.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157:1819-1829.
- Ogawa, Sh., Matsuda, H., Taniguchi, Y., Watanabe, T., Sugimoto, Y. and Iwaisaki, H. (2016). Estimated Genetic Variance Explained by Single Nucleotide Polymorphisms of Different Minor Allele Frequencies for Carcass Traits in Japanese Black Cattle. *Journal of Biosciences and Medicines*. 4: 89-97.
- تیموریان، م.، اسلمی نژاد، ع.ا. و شریعتی، م.م. (۱۳۹۴). بررسی اثر برخی عوامل جمعیتی بر کیفیت انتخاب ژنومیک در گاوهای هلشتاین ایران. نشریه پژوهشهای علوم دامی ایران. جلد ۷، شماره ۱، صص ۹۶-۱۰۳.
- دیمی غیاث آبادی، پ.، علیجانی، ص.، شجاع غیاث، ج. و پیرانی، ن. (۱۳۹۱). مقایسه دو روش آماری حداکثر درست نمایی محدود شده و آنالیز بیزی در برآورد پارامترهای ژنتیکی برخی صفات مهم اقتصادی مرغهای بومی استان فارس. پژوهشهای تولیدات دامی. سال سوم. شماره ۵. صص ۱۳-۱.
- راشدی ده صحرائی، آ.، فیاضی، ج.، وطن خواه، م. و بیگی نصیری، م.ت. (۱۳۹۲). برآورد اجزای (کو) واریانس و فراسنجه های ژنتیکی صفات رشد در برهه های لری بختیاری با استفاده از روش نمونه گیری گیسس. نشریه پژوهش در نشخوارکنندگان. جلد ۱. شماره ۲. صص ۱۲۸-۱۰۹.
- قائم، م.، افشار، م.، عسگری جعفرآبادی، ق. و حاجی بنده، ن. (۱۳۹۱). اثر ضریب وراثت پذیری بر صحت ارزیابی ژنومی و تغییرات ضریب همخونی. مجله دانش و پژوهش علوم دامی / جلد ۱۰، صص ۹۵-۱۰۲.
- Abdollahi-Arpanahi, R., Pakdel, A., Nejati-Javaremi, A., Moradi Shahrababak, M., Morota, G., Valente, B.D. et al. (2014). Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. *Animal Breeding and Genetics*. 131: 183-193.
- 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.
- Gianola, D. and de los Campos, G. (2013). Genome-enabled prediction of complex traits. University of Wisconsin-Medison, May 27th-31st, 2013.
- Hayes, B.J. (2007). QTL, Mapping Mas, and Genomic selection Animal breeding and Genetics Department of Animal Science

