

تعیین معادله تابعیت صفت وزن شتر از صفات بیومتری شتران نژاد یزدی با استفاده از آنالیز رگرسیون خطی چند متغیره به روش تحلیل مؤلفه های اصلی

- حسین مرادی شهر بابک (نویسنده مسئول)

استاد یار گروه مهندسی علوم دامی دانشکده مهندسی کشاورزی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران

- هادی مقبلی

دانش آموخته کارشناسی ارشد ژنتیک و اصلاح نژاد، پردیس کشاورزی و منابع طبیعی دانشگاه تهران، دانشکده مهندسی کشاورزی، گروه مهندسی علوم دامی

- محمد مرادی شهر بابک

استاد گروه مهندسی علوم دامی دانشکده مهندسی کشاورزی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران

تاریخ دریافت: مرداد ۱۳۹۲ تاریخ پذیرش: دی ۱۳۹۲

شماره تماس نویسنده مسئول: ۰۲۶۳۲۲۴۸۰۸۲

Email: hmoradis@ut.ac.ir

- سید رضا میرایی آشتیانی

استاد گروه مهندسی علوم دامی دانشکده مهندسی کشاورزی، پردیس کشاورزی و منابع طبیعی دانشگاه تهران

چکیده

هدف از این مطالعه تعیین معادله تابعیت وزن زنده از چهار صفت طول بدن، محیط دور سینه، ارتفاع جدوگاه و محیط شکم در ۱۵۰ نفر از شترهای یک کوهانه نژاد یزدی بود. سعی شد تا جزئیات مبسوطی درباره روشی مناسب جهت کاهش مشکلات هم خطی و هم بستگی بین متغیرهای مستقل و برآورد وزن زنده شتر بیان شود. نتایج حاصل از این مطالعه نشان دادند که مشکل هم راستایی چندگانه موجود در اطلاعات مربوط به ارتباط بین وزن با متغیرهای مستقل مربوطه، با استفاده از روش تابعیت مؤلفه های اصلی قابل حل است. از بین متغیرها، صفت ارتفاع بدن بیشترین و صفت دور شکم دارای کمترین ضریب در برآورد وزن زنده بودند. پیشنهاد می شود که در صورت وجود هم راستایی در آنالیزهای رگرسیون خطی چند متغیره، روش آنالیز مؤلفه های اصلی به کار برده شود که برآوردهای دقیق تری نسبت به روش های حداقل مربعات معمولی دارد؛ همچنین، این روش می تواند با برآورد دقیق بعضی از صفات مهم که اندازه گیری آن ها مشکل است و انتخاب بهترین متغیرها برای برآورد صفات در انجام انتخاب به اصلاح گران کمک کند.

واژه های کلیدی: شتر یک کوهانه، هم راستایی چندگانه، تحلیل مؤلفه های اصلی.

Animal Science Journal (Pajouhesh & Sazandegi) No 108 pp: 25-34

Determination of camel weight regression equation using biometrical traits of Yazdi camels bred by multivariate linear regression analysis based on the principal components analysis.

Hossein Moradi Shahrabak^{*1}, Hadi Moghbeli, Mohammad Moradi Shahrabak³ and Seyed Reza Miraei Ashtiani³

1. Assistant professor of Department of Animal Science, Faculty of Agricultural Sciences and Engineering, University College of Agriculture & Natural Resources – University of Tehran, Karaj, Iran.

2. Master Science University of Tehran, Department of Animal Science, Faculty of Agricultural Sciences and Engineering, University College of Agriculture & Natural Resources – University of Tehran, Karaj, Iran.

3-professor of Department of Animal Science, Faculty of Agricultural Sciences and Engineering, University College of Agriculture & Natural Resources – University of Tehran, Karaj, Iran

Corresponding Author E-Mail: hmoradis@ut.ac.ir

Received: July 2013

Accepted: December 2013

The objective of the current study was to determine the live body weight regression equation using four traits including body length, heart circumference, height of withers and abdominal circumference in 150 camels of Yazdi breed. It was tried to describe the adequate details about an appropriate method for fixing or decreasing the problems of collinearity instability and correlation between independent variables related to live body weight of Yazdi camel ecotype. The results of this study showed that the problem of multicollinearity within the data was associated to body weight of Yazdi camel breed and related independent variables could be fixed by the principal components analysis. Among the independent variables, body height and abdominal circumference had the highest and lowest coefficient in estimating the body weight of Yazdi camel, respectively. It suggested that the principal components analysis be used when there is the problem of collinearity in multivariable linearly regression analysis because of more precise estimation than least squares method. Also this method can help to the breeders to select the best animals by predicting the precise value of some important traits and selecting the best independent variables for predicting the traits.

Key words: Yazdi camel, Multicollinearity, Principal Components Analysis.

مقدمه

توجه خاصی دارد. به منظور حفظ و بهبود عملکرد این اکوتیپ برای تعیین ویژگی‌های جمعیت‌ها، انتخاب و اصلاح نژاد پیوسته مورد نیاز می‌باشد و اجرای برنامه‌های اصلاحی، ضرورت خاصی دارد (Khatami, 1984). اولین مرحله در کارهای اصلاحی و انتخاب حیوان، شناسایی صفات اقتصادی مهم و اندازه‌گیری و یا برآورد دقیق صفات می‌باشد و استفاده از روش‌های مناسب و صحیح در این خصوص ضروری است. انتخاب و تعیین صحیح صفات مورد نظر، اولین و مهمترین اولویت در اجرای هر برنامه اصلاح نژادی است. به دلیل نبود اطلاعات کافی از شتران ایرانی، وجود نداشتن گله‌های مولد مناسب و به دلیل موقعیت و شرایط مکانی زندگی شترها، در اکثر موارد دسترسی به باسکول امکان پذیر نیست و اندازه‌گیری وزن بدن شتر با مشکلاتی همراه است. لذا اهمیت داشتن معادله صحیحی از تابعیت وزن از صفات

در ایران سه گروه شتر دو کوهانه، یک کوهانه و آمیخته وجود دارند، که هر کدام شامل اکوتیپ‌های متعددی نیز هستند. منابع ژنتیکی شتر در ایران بیشتر از نوع شتر یک کوهانه (*Camelus dromedarius*) است و شترهای دو کوهانه (*Camelus bactrianus*) که به تعداد خیلی کم و در حال انقراض بوده و در استان اردبیل وجود دارند. با توجه به این که گونه‌های شتر دارای تولیدات و خدمات فراوان شامل پشم، گوشت، شیر و حمل و نقل می‌باشند، حفظ تنوع ژنتیکی شتر و اصلاح نژاد آن دارای اهمیت بالایی است. زیرا نه فقط برای پیشرفت ژنتیکی و مدیریت این منابع لازم است، بلکه برای پایداری سیستم‌های تولیدی و قابل دسترس بودن تنوع تولیدات دامی نیز حائز اهمیت است. در این میان، شتر نژاد یزدی نیز از نظر آسیب‌های جمعیتی و آسیب‌های ناشی از خشک سالی‌های پی در پی در حوضه استان یزد نیاز به

(*et al.*, 2008). رش آنالیز مؤلفه های اصلی^۳ یک روش ریاضی جهت تبدیل تعدادی از متغیرهای همبسته به تعداد کمتری از متغیرهای غیرهمبسته که مؤلفه های اصلی نامیده می شوند، می-باشد (Freund and Littell, 2000; Karacaoren and Kadarmideen, 2008) از نقطه نظر ژنتیک و اصلاح حیوان، مؤلفه های اصلی، گروهی از صفاتی که می تواند به منظور اهداف انتخابی استفاده شود را مشخص می کند (Sangdon, 2006). در واقع، با استفاده از آنالیز مؤلفه های اصلی قادر هستیم در صورت اندازه گیری یکسری از صفات مرتبط با صفت مورد نظر، اندازه دقیق صفت مورد نظر را پیش بینی کنیم. فریتز و همکاران (۱۹۷۱) برای اولین بار روش تابعیت مؤلفه های اصلی را برای برآورد ضرایب تابعیت در داده های با مشکل هم راستایی چند گانه ارائه کردند (Fritts *et al.*, 1971). محققان زیادی از روش آنالیز مؤلفه های اصلی جهت تخمین وزن بدن و صفات عملکردی استفاده کرده اند (Yugfin, 2008). آنالیز مؤلفه های اصلی جهت توصیف همبستگی بین اندازه گیری های ظاهری مربوط به بدن و اندازه بدن در جوجه های گوشتی (Yakubu and Idahor, 2009) و بوقلمون (Ogah, 2011) استفاده شده است. پینتو و همکاران (۲۰۰۶)، از این روش برای تعیین معیار انتخاب برای بهبود اندازه بدن استفاده کردند (Pinto and Dacker, 2006). این تکنیک همچنین برای کاهش تعداد متغیرهای مستقل جهت پیش بینی ارزش های اصلاحی ژنومیکس استفاده شده است (Macciotta and Gaspa, 2010). هدف مطالعه حاضر ارائه روش مناسب، ساده و دقیق و کارآمد جهت برآورد وزن زنده شتر یک کوهانه اکوتیپ یزدی و استفاده از این اطلاعات در ارزیابی شتران این نژاد است.

مواد و روش ها

این تحقیق، با استفاده از رکوردهای اندازه گیری شده از صفات بیومتری و اندازه گیری وزن بر روی ۱۵۰ نفر از شتران نژاد یزدی انجام گرفت. شتران مورد آزمایش در این مطالعه از مناطق مختلف استان یزد و در گروه های سنی مختلف به صورت تصادفی انتخاب گردیدند. جنسیت، سن، وضعیت آبستنی، وزن، اندازه های بدن در

بیومتری بیش از پیش نمایان می شود و ضرورت ارائه روش مناسب در این خصوص سبب گردیده است تا با استفاده از اندازه های ظاهری بدن، وزن شتر تخمین زده شود.

زمانی که بین متغیرهای پیش بینی کننده همبستگی وجود دارد، گفته می شود حالت هم راستایی چند گانه^۱ وجود دارد. در صورتی که متغیرهای پیش بینی کننده همبستگی بالایی داشته باشند، خطای استاندارد ضرایب تابعیت جزئی برآورد شده زیاد بوده (به عبارتی واریانس برآوردها را افزایش می دهد) و تفسیر آن ها نیز غیر ممکن است. بنابراین، تفسیر ساده ضرایب تابعیت جزئی به طوری که اثرات عمده را اندازه گیری کنند، تضمین نشده می باشد. هم-راستایی چند گانه مشکلی است که در همه جا حاضر بوده و بر روی فرض معکوس ماتریس تاثیر می گذارد (Lee, 2006).

در نتیجه، با وجود هم راستایی چند گانه در داده ها نمی توان از روش های تابعیت حداقل مربعات معمولی^۲ استفاده کرد و این مشکل در چهار خصوصیت مدل شامل مقدار، علامت، خطاهای استاندارد و ضریب تبیین که برای ما بسیار مهم می باشد، تاثیر منفی می گذارد (Motulsky, 2009; Guiot *et al.*, 1982).

اگر بین متغیرهای مستقل در تابعیت چند متغیره رابطه خطی وجود داشته باشد، گفته می شود که نامتعامل هستند. وقتی متغیرهای مستقل متعامل باشند، به آسانی می توان از تابعیت چند متغیره استفاده کرد. مکانیسم های درگیر در کنترل اغلب صفات بیولوژیکی در موجودات زنده پیچیده تر از آن هستند که به وسیله آنالیز تک متغیره تفسیر شوند زیرا اغلب صفات به طور بیولوژیکی از طریق اثرات پلوتروویک ژن ها و پیوستگی جایگاه های ژنی با همدیگر همبستگی دارند (Udeh and Ogbu, 2011). متأسفانه در اکثر موارد متغیرهای مستقل متعامل نیستند. گاهی متعامل نبودن، مشکلاتی را فراهم نمی سازد اما در برخی موارد، متغیرهای مستقل همبستگی دارند و در چنین مواردی استنتاج بر مبنای مدل تابعیت می تواند گمراه کننده باشد. در حالتی که بین متغیرهای مستقل وابستگی خطی مشاهده شود، گفته می شود که هم راستایی چند گانه وجود دارد (Rosario Loehlin, 1998);

¹ Multicollinearity

² Ordinary Least Squares

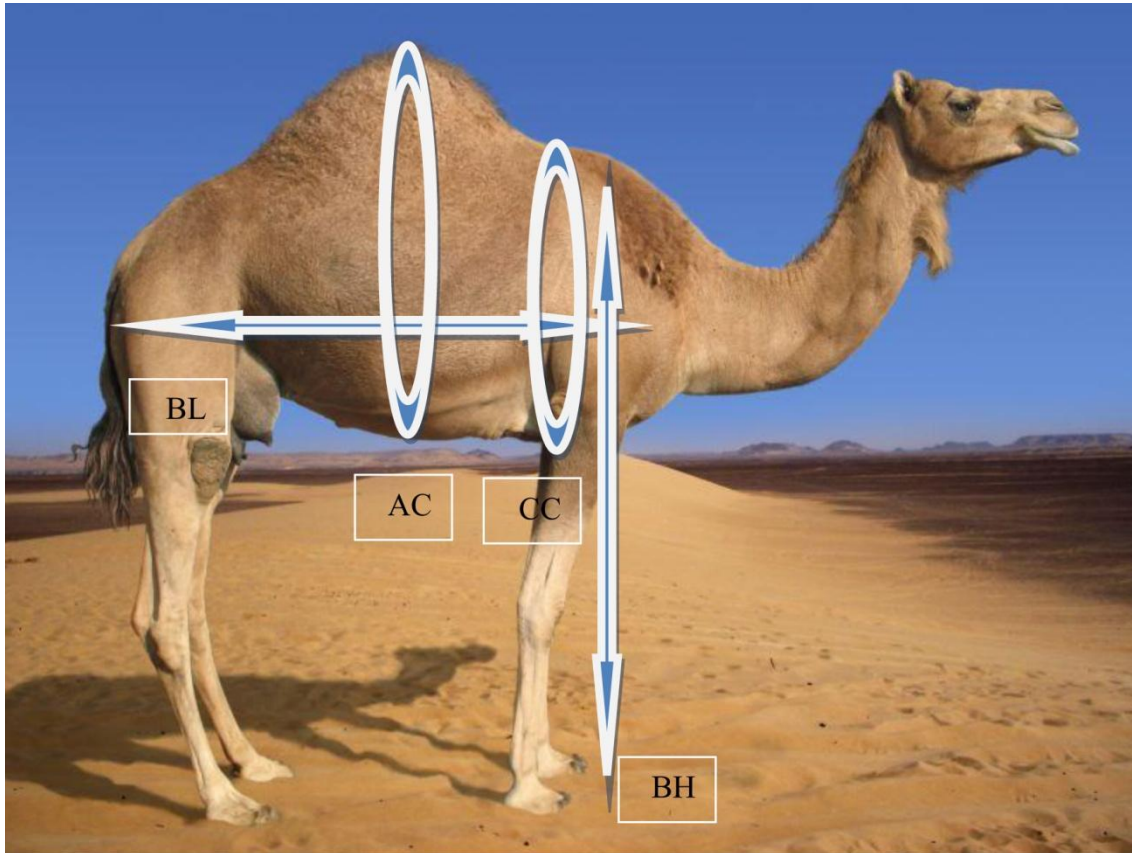
³ Principle Component Analysis (PCA)

گونه خمیدگی؛ طول بدن (از سر شانه تا انتهای ران)؛ دور سینه (از پشت ارنج تا روی دنده جلوی کوهان تا پشت دستها، دور کامل)؛ محیط شکم (از بالای بلندترین نقطه کوهان از ناحیه زیر شکم، دور کامل) بود که محل اندازه گیری آن در شکل ۱ آورده شده است.

فرم‌های مربوطه یادداشت شد و صفات بیومتری با استفاده از متر پارچه‌ای و کولیس با دقت یک سانتیمتر اندازه‌گیری و ثبت شدند.

صفات بیومتری شتر و نحوه اندازه‌گیری

صفات مورد بررسی شامل ارتفاع بدن از ناحیه جدوگاه (به موازات دستهای شتر تا سطح زمین در حالت ایستاده کامل و عاری از هر



شکل ۱- اندازه‌گیری صفات بیومتری شده (BH: ارتفاع بدن از جدوگاه، BL: طول بدن، CC: دور سینه، AC: محیط شکم)

مدل به کار رفته در این مورد در زیر نشان داده شده است.

$$y_i = b_0 + b_{1i}(x_1) + b_{2i}(x_2) + b_{3i}(x_3) + b_{4i}(x_4) + e_i$$

Y متغیر وابسته (در این جا صفت وزن زنده)، b_0 عرض از مبدأ، Xها متغیرهای مستقل (در این جا ۴ صفت مورد استفاده برای پیش-بینی صفت وزن)، b_1 - b_4 ضرایب تابعیت متغیرهای مستقل که باید برآورد شوند، e خطای باقیمانده بوده که فرض می‌شود با مشاهدات همبستگی نداشته و میانگین صفر و واریانس ثابت دارند.

پس از ورود داده‌ها به رایانه، جهت تعیین بهترین معادله تابعیت آنالیز صورت گرفت. متغیرهای مستقل عبارت بودند از: ۱- ارتفاع بدن از جدوگاه، ۲- طول بدن حیوان، ۳- محیط دور شکم، ۴- محیط دور سینه.

برای انجام تابعیت چندگانه از نرم‌افزارهای SAS 9.1 رویه GLM و همچنین EXCEL EX.SAT استفاده شد که هر دو نتایجی مشابه ارائه کردند.

مقدار می‌باشد که از طریق زیر قابل دست یابی می‌باشد:

$$\frac{\sum_{j=1}^r \lambda_j}{k} > 0.99$$

λ = مقادیر ویژه ماتریس همبستگی

k = تعداد بردارهای ویژه نرمال شده

r = تعداد بردارهای ویژه نرمال مورد نظر

j = تعداد مقدار ویژه مورد نظر

این مقادیر با استفاده از روش PCA (Proc Princomp) در نرم افزار SAS 9.1 برآورد می‌شوند که هر بردار ویژه در به صورت مولفه های اصلی^۴ به تعداد متغیرهای مستقل موجود در مدل ارائه شده و مقادیر ویژه لازم انتخاب شدند و برای ایجاد معادله تابعیت جدید مورد استفاده قرار گرفتند. معادله برآورد ضرایب هر کدام از بردارهای ویژه مورد استفاده در زیر نشان داده شده است.

$$y_i = B_0 + B_{1i}(prin_1) + \dots + B_{ki}(prin_k) + e_i$$

که پیرین ها در این معادله برابر با بردارهای ویژه مربوط به مقدار ویژه مربوطه می‌باشد. λ برابر با متغیر وابسته استاندارد شده بوده و B ها ضرایب تابعیت مربوطه می‌باشند (B_0 نیز عرض از مبدا می‌باشد). پس از برآورد ضرایب تابعیت مولفه های اصلی با کاربرد معادله ماتریسی زیر ضرایب تابعیت متغیرهای مستقل استاندارد شده (برآورد کننده‌های مولفه‌های اصلی) به دست آمد (Loehlin, 1998; Jolliffe, 2005).

$$\begin{bmatrix} b_{1,pc} \\ \vdots \\ b_{k,pc} \end{bmatrix} = \begin{bmatrix} prin_1 & \dots & prin_k \end{bmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_k \end{bmatrix}$$

در این معادله b_{pc} ها برابر با ضرایب تابعیت متغیرهای مستقل استاندارد شده می‌باشند. برای بدست آوردن ضرایب تابعیت متغیرهای مستقل قبل از استاندارد کردن از فرمول

به منظور بررسی وجود هم‌راستایی چندگانه در داده‌های مورد بررسی روش‌های مختلف مورد آنالیز قرار گرفتند. روش اول بررسی همبستگی بین داده‌ها می‌باشد که همبستگی بین داده‌ها از طریق نرم افزار SAS 9.1 رویه Proc corr محاسبه شد که همبستگی بالا بین متغیرهای مستقل از نشانه‌های هم‌راستایی چندگانه است. همچنین Variance Inflation Factor (VIF) و شاخص وضعیت نیز مورد بررسی قرار گرفت.

$$tolerance = 1 - R_j^2, \quad VIF = \frac{1}{tolerance}$$

در این دو معیار، ضریب تعیین رگرسیون متغیر توضیح دهنده R_j^2 از روی دیگر متغیرهای توضیح دهنده است. چنانچه tolerance کوچکتر از ۰/۲ یا VIF بزرگتر از ۱۰ باشد در آن صورت هم‌راستایی محتمل است.

در ادامه از رویه GLM و PRINCOMP در نرم‌افزار SAS مورد استفاده قرار گرفت. به منظور انجام روش تابعیت مولفه‌های

$$x^{**} = \frac{x_{ij} - \bar{x}_i}{s_i}$$

استاندارد شدند. به طوری که x^* متغیر مستقل استاندارد شده بوده و \bar{x}_i و s_i به ترتیب میانگین و انحراف معیار متغیر مستقل مربوطه می‌باشند. بعد از استاندارد کردن و نرمال سازی داده‌ها، از رویه PRINCOMP نرم‌افزار SAS برای محاسبه مقادیر ویژه^۴ و بردارهای ویژه^۵ مربوط به هر کدام از مقادیر ویژه استفاده شد. پس از برآورد مقادیر و بردارهای ویژه مربوطه با استفاده از استراتژی حذف مولفه‌های اصلی حذف مولفه‌ها با کمترین مقدار ویژه شروع می‌شود.

این منطقی است که مولفه اصلی با کمترین مقدار ویژه کمترین اطلاعات را در بر دارد. با استفاده از این روش، مولفه‌های اصلی حذف می‌شوند تا این که مولفه‌های باقیمانده درصدی از کل واریانس را که از قبل تعیین کردیم توضیح دهند.

در حقیقت این روش انتخاب مجموعه‌ای از مقادیر ویژه با بیشترین

⁴ Eigenvalue

⁵ Eigenvector

⁶ Prin

بر آورد کننده‌های مولفه اصلی از ضرایب تابعیت در نهایت با معادله $b_{j.pc}^* = \frac{b_{j.pc}^0}{s_j}$ که s_j برابر با خطای استاندارد متغیر مستقل مربوطه می‌باشد، به دست می‌آید. در نتیجه خطای استاندارد بر آورد کننده‌های مولفه اصلی برای ضرایب متغیرهای مستقل اولیه (قبل از استاندارد کردن) به این صورت محاسبه می‌شود:

$$s.e.(b_{j.pc}^*) = \sqrt{\text{var}\left(\frac{b_{j.pc}^0}{s_j}\right)}$$

که در این معادله $s.e.(b_{j.pc}^*)$ خطای استاندارد بر آورد کننده مولفه اصلی از ضریب مربوطه به z امین متغیر مستقل می‌باشد. یکی از معیارهای مهم برای مقایسه این دو روش خطای استاندارد بر آوردها می‌باشد زیرا همان‌طور که بیان شد یکی از پیامدهای هم‌راستایی چندگانه تورم واریانس بر آوردها می‌باشد، در نتیجه بر آوردهای به دست آمده از روش تابعیت مولفه‌های اصلی صحت بالاتری داشته و خطای استاندارد کمتری خواهند داشت (Gunst ; Jeeshim, 2002; Jolliffe *et al.*, and Mason, 1980 ; 2003; Smith, 2002).

نتایج و بحث

ضریب تعیین این بر آوردها ۰/۹۴ و ضریب تغییرات وزن بر آورد شده برابر با ۲۹ بود. در جدول یک، خلاصه آماری صفات مورد استفاده نشان داده شده است.

جدول ۱- خلاصه آماری صفات اندازه‌گیری شده

ضریب تغییرات	میانگین	انحراف معیار	بیشینه	کمینه	صفت
۰/۴۷	۲۳۵/۶۴	۱۱۱/۰۴	۴۱۲	۷۹	وزن بدن (kg)
۰/۱۷	۱۲۱/۵۰	۲۱	۱۶۲	۸۷	ارتفاع (cm)
۰/۱۱	۱۵۳/۶۲	۱۷/۵۴	۱۹۸	۱۲۸	طول بدن (cm)
۰/۱۵	۱۹۱/۳۵	۲۹/۲۳	۲۵۸	۱۳۸	محیط شکم (cm)
۰/۱۱	۱۶۶/۱۱	۱۷/۵۴	۲۰۲	۱۳۱	محیط سینه (cm)

$$b_{j.pc}^* = \frac{b_{j.pc}^0}{s_j}, j = 1, 2, 3, 4$$

ضریب تابعیت متغیر مستقل قبل از استاندارد کردن می‌باشد. همچنین برای محاسبه عرض از مبدا از معادله زیر استفاده شد.

$$b_{0.pc}^* = b_{0.pc} - \frac{b_{1.pc} x_1}{s_1} - \frac{b_{2.pc} x_2}{s_2} - \frac{b_{3.pc} x_3}{s_3} - \frac{b_{4.pc} x_4}{s_4}$$

که $b_{0.pc}^*$ برابر با عرض از مبدا برای معادله تابعیت متغیرهای مستقل قبل از استاندارد کردن می‌باشد. برای محاسبه خطاهای استاندارد این بر آوردها نیز از معادلات زیر استفاده شد (RH, 1986; Waring *et al.*, Hocking, 2005) (2010).

$$\text{var}(b_{pc}^*) = \begin{bmatrix} \text{var}(B_1) \\ \vdots \\ \text{var}(B_4) \end{bmatrix} \begin{bmatrix} \text{prin}_{11}^2 & \dots & \text{prin}_{1k}^2 \\ \vdots & \ddots & \vdots \\ \text{prin}_{41}^2 & \dots & \text{prin}_{4k}^2 \end{bmatrix}$$

در نتیجه $\text{var}(b_{pc}^*)$ برابر است با واریانس بر آورد کننده‌های مولفه اصلی از ضرایب متغیرهای مستقل که استاندارد شده‌اند. خطای استاندارد b_{pc}^* نیز به این صورت محاسبه می‌شود:

$$s.e.(b_{pc}^*) = \left[\text{var}(b_{pc}^*) \right]^{1/2}$$

اصلی استفاده کنیم. در این صورت برآوردهای دقیق تر و با خطای استاندارد کمتر به دست خواهد آمد.

جدول ۳- مقادیر تورم واریانس برای متغیرهای مستقل

تورم واریانس	صفت
۵۰/۳۵	ارتفاع
۰/۴۸	طول بدن
۱/۳۹	محیط شکم
۲/۸۲	دور سینه به توان ۳

مرحله بعد، محاسبه معادله تابعیت مورد نظر از طریق روش تابعیت مولفه‌های اصلی می‌باشد. بدین منظور، ابتدا باید مقادیر ویژه و بردارهای ویژه مربوطه را برآورد کرد. در جدول (۴) و (۵) به ترتیب مقادیر ویژه و بردارهای ویژه برآورد شده نشان داده شده است. همان‌طور که در جدول (۴) مشاهده می‌شود، به ترتیب مقادیر ویژه از بالا به پایین کاهش می‌یابند. همچنین با افزایش عدد مقادیر ویژه، میزان سهم آن‌ها در توضیح واریانس نیز کاهش می‌یابد. به این ترتیب، مقدار ویژه اول بالاترین سهم از کل واریانس ۰/۹۱۶۵ و مقدار ویژه آخر نیز کمترین مقدار از کل واریانس ۰/۰۱۱۷ را توضیح می‌دهند. مقدار ویژه دوم، بخشی از واریانس کل را که مقدار ویژه اول توضیح نداده، توضیح می‌دهد و از طرف دیگر، نشان می‌دهد که با مقدار ویژه اول همبسته نیست. این دو خصوصیت برای مقادیر ویژه بعدی وجود دارند. در نتیجه با استفاده از تابعیت مولفه‌های اصلی، داده را متعامد یا غیر همبسته کرده و سپس معادله تابعیت مورد نظر به دست می‌آید.

جدول ۴- مقادیر ویژه برای متغیرهای مورد استفاده

مقادیر ویژه	واریانس*	ردیف
۳/۶۷	۰/۹۱۶۵	۱
۰/۱۷	۰/۰۴۲۲	۲
۰/۱۲	۰/۰۲۹۵	۳
۰/۰۴۷	۰/۰۱۱۷	۴

* منظور از واریانس میزان واریانس است که این مقدار ویژه از کل واریانس موجود توضیح می‌دهد.

در این بررسی برای بدست آوردن تابعیت به روش حداقل مربعات، با ضریب تبیین بالایی روبرو خواهیم شد که به دلیل همبستگی بالای صفات می‌باشد و این حالت می‌تواند یکی از ابتدایی‌ترین نشانه‌های وجود هم‌راستایی چندگانه باشد. برای این منظور قدم بعدی کنترل همبستگی‌ها بین متغیرهای مستقل است. در صورت وجود همبستگی بالا باید معیارهای معتبرتر دیگری را برای تایید وجود هم‌راستایی چندگانه به کار برد. در جدول ۲ همبستگی‌های بین متغیرهای مستقل ارائه شده است.

با مشاهده همبستگی‌های بین متغیرها متوجه خواهیم شد که بین اکثر متغیرها همبستگی بالایی وجود دارد. برآورد ضریب تبیین بالای به دست آمده در این حالت، ناشی از تورم واریانس حاصل از هم‌راستایی بوده که ناشی از هم‌راستایی چندگانه است. در این حالت شرط نبود رابطه بین متغیرها برای معکوس ماتریس $X'X$ به منظور برآورد پارامترها در روش حداقل مربعات معمولی به هم خورده و تورم واریانس ایجاد می‌شود که با وجود خطای استاندارد بالا برای بعضی از برآوردها نیاز به یک روش بهتر برای برآوردهای دقیق تر می‌باشد. برای بررسی هم‌راستایی چندگانه عامل تورم واریانس نیز محاسبه شد. در جدول ۳ این مقدار برای متغیرها نشان داده شده است.

جدول ۲- همبستگی‌های بین متغیرهای مستقل

صفت	(۱)	(۲)	(۳)	(۴)
ارتفاع (۱)	۱	۰/۹۷**	۰/۹۴**	۰/۹۵**
طول بدن (۲)	-	۱	۰/۸۹**	۰/۹۳**
محیط شکم (۳)	-	-	۱	۰/۹۵**
محیط دور سینه (۴)	-	-	-	۱

** معنی دار در سطح ($p < 0.01$)

با توجه به جدول ۳ مشاهده می‌گردد که صفت ارتفاع، واریانس بیشتری از مقدار بحرانی (مقدار بحرانی ۵ تا ۱۰ می‌باشد) دارد و در این حالت هم‌راستایی چندگانه رخ خواهد داد. در صورتی که نخواهیم این متغیرها را از مدل حذف کنیم و حضور آن‌ها در مدل ضروری باشد، باید از روش آماری مناسبی مانند تابعیت مولفه‌های

در نتیجه به عنوان مثال مقدار ویژه اول برابر با $3/67$ می‌باشد، به عبارت دیگر این مقدار ویژه $3/67$ واحد از ۴ واحد واریانس را توضیح می‌دهد و به همین ترتیب برای دیگر مقادیر ویژه نیز به همین ترتیب برآورد می‌شود.

با توجه به این که داده‌ها در این روش استاندارد می‌شوند و دارای میانگین صفر و واریانس یک می‌باشند، پس واریانس کل برابر با تعداد متغیرهایی است که در آنالیز شرکت می‌کنند که در این مطالعه برابر با ۴ است.

جدول ۵- بردارهای ویژه برای متغیرهای مورد استفاده

مولفه های اصلی	۱	۲	۳	۴
ارتفاع	۰/۴۹	۰/۷۸	-۰/۳۹	۰/۰۶
طول بدن	۰/۵۰	۰/۱۲	۰/۸۵	-۰/۰۸
محیط شکم	۰/۵۱	-۰/۴۶	-۰/۱۶	۰/۷۱
محیط دور سینه	۰/۵۱	-۰/۴۱	-۰/۳۰	-۰/۶۹

جدول ۶- برآوردهای به دست آمده برای معادله تابعیت با استفاده از تابعیت مولفه‌های اصلی

ردیف	برآورد	خطای استاندارد
عرض از مبدا	۲۳۹/۷	۴/۶۸
مولفه ۱	۵۷/۰۴	۲/۴۴
مولفه ۲	۰/۱۷	۱۱/۴
مولفه ۳	۰/۲۷۹	۱۳/۶۲
مولفه ۴	-۳/۱۹۸	۲۱/۶۱

ارائه ضرایب تابعیت متغیرهای مستقل استاندارد شده برای تحقیقات آینده روی دام‌هایی با شرایط متفاوت (مثل دام‌هایی که در شرایط کشور ایران پرورش نیافته‌اند)، نسبت به دام‌های مورد مطالعه در تحقیق حاضر، مفید است.

نکته مهم در استفاده از ضرایب جدول ۷ در تحقیقات بعدی این است که این ضرایب حتماً بایستی در مقادیر استاندارد شده متغیرهای مستقل ضرب شوند.

همان‌طور که بیان شد، وجود تعداد کمی از مقادیر ویژه با مقدار زیاد نشان می‌دهد که تعداد کمی از متغیرها بیشتر تنوع موجود در متغیرهای مشاهده شده را توضیح می‌دهند. مقدار ویژه صفر یعنی هم‌راستایی کامل بین متغیرهای مستقل بوده و مقادیر ویژه بسیار کوچک نشان دهنده هم‌راستایی چندگانه قوی می‌باشد. با توجه به جدول ۵ نیز مشاهده می‌گردد که تقریباً مقدار ویژه اول بیشترین واریانس را توضیح داده و بقیه مقادیر ویژه نقش کمتری را دارند که این دلیلی دیگر بر وجود هم‌راستایی چندگانه می‌باشد. برای انتخاب تعداد مقادیر ویژه مورد نیاز انجام تابعیت مولفه‌های اصلی،

از فرمول $\frac{\sum_{j=1}^r \lambda_j}{k} > 0.99$ استفاده شد و در نتیجه مولفه ۱ از ۴ مولفه موجود مورد استفاده قرار گرفت. در جدول ۶ این برآوردها به همراه خطاهای استانداردشان نشان داده شده است. سپس به منظور تبدیل این برآوردها به مقادیر اولیه ابتدا این مقادیر باید به ضرایب تابعیت متغیرهای مستقل استاندارد شده (برآورد کننده‌های مولفه‌های اصلی) و سپس به ضرایب تابعیت برای متغیرهای مستقل قبل از استاندارد شدن، تبدیل گردند. در جدول ۷ نتیجه این برآوردها به همراه خطاهای استاندارد مربوطه نشان داده شده است.

روش حداقل مربعات معمولی استفاده کرد که در این صورت، خطاهای استاندارد برآوردها نسبت به روش حداقل مربعات معمولی، کمتر خواهد بود. بر اساس نتایج به دست آمده در این تحقیق پیشنهاد می شود که از آنالیز مؤلفه‌های اصلی در شرایط وجود هم‌راستایی در آنالیزهای رگرسیون خطی چند متغیره و برای کاهش مؤثر در تعداد صفات بیومتريک مورد نیاز برای استفاده در کارهای اصلاحی و انتخاب‌ها استفاده شود. به همین علت، مشکل هم‌راستایی چندگانه موجود در اطلاعات مربوط به ارتباط بین وزن شتر نژاد یزدی با ۴ متغیر مستقل مربوط به این صفت با استفاده از روش تابعیت مؤلفه‌های اصلی قابل حل می‌باشد.

منابع

- Freund, R. J. & Littell, R. C. (2000). *SAS system for regression*. SAS Institute.
- Fritts, H. C., Blasing, T. J., Hayden, B. P. & Kutzbach, J. E. (1971). Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate. *Journal of applied meteorology* 10(5): 845-864.
- Guiot, J., Berger, A., Munaut, A. & Till, C. (1982). Some new mathematical procedures in dendroclimatology, with examples from Switzerland and Morocco. *Tree-Ring Bulletin* 42: 33-48.
- Gunst, R. F. & Mason, R. L. (1980). *Regression analysis and its application: a data-oriented approach*. CRC Press.
- Hocking, R. R. (2005). *Methods and applications of linear models: regression and the analysis of variance*. Wiley. com.
- Jeeshim, T. (2002). Multicollinearity in Regression Models. <http://www.masil.org/documents/multicollinearity.pdf>. 31: 10-19.
- Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library.
- Jolliffe, I. T., Trendafilov, N. T. & Uddin, M. (2003). A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics* 12(3): 531-547.

جدول ۷- ضرایب تابعیت برآورده شده برای متغیرهای مستقل استاندارد شده و خطای استاندارد برآوردها با استفاده از روش تحلیل مؤلفه‌های اصلی

صفت	برآورد	خطای استاندارد
ارتفاع	۰/۲۴۶	۰/۰۹۳
طول بدن	۰/۲۵۷	۰/۱۰۶
محیط شکم	۰/۲۳۵	۰/۱۴۶
دور سینه	۰/۲۷۴	۰/۱۴۵

جدول ۸- برآوردهای به دست آمده برای معادله تابعیت بعد از تبدیل برای متغیرهای ابتدایی

صفت	برآورد	خطای استاندارد
عرض از مبدا	-۵۷۳/۳۲	۵۱/۳۶
ارتفاع	۱/۵۹	۰/۶۲
طول بدن	۱/۳۷	۰/۵۶
محیط شکم	۰/۹۰	۰/۵۶
محیط سینه	۱/۳۷	۰/۷۲

با توجه به برآوردهای صورت گرفته در این مطالعه بر اساس روش تابعیت مؤلفه‌های اصلی، در معادله تابعیت صفت ارتفاع بیشترین و صفت محیط شکم، کمترین ضریب را در برآورد وزن نشان دادند. مالووا و همکاران (۲۰۰۴) در مطالعه‌ای برای صفت وزن بدن در گاوهای سیاه ژاپنی، برآورد معادله تابعیت را با روش‌های تابعیت ریح و مؤلفه‌های اصلی با استفاده از ۱۳ متغیر مستقل بررسی کردند. بر اساس نتایج آن‌ها، استفاده از آنالیز مؤلفه‌های اصلی برای برآورد ضرایب تابعیت پایدار و به منظور غلبه بر مشکل هم‌راستایی چندگانه مناسب می‌باشد (Malau-Aduli et al., 2004).

نتیجه گیری

با توجه به وجود هم‌راستایی چندگانه در داده‌ها، برآوردهای به دست آمده از روش حداقل مربعات معمولی ممکن است با تغییر داده‌ها (کم یا زیاد شدن داده‌ها) تغییر کند و در نتیجه معتبر نمی‌باشد. در صورت وجود هم‌راستایی چندگانه در داده‌ها باید از روش تابعیت مؤلفه‌های اصلی برای برآورد معادله تابعیت نسبت به

- Karacaoren, B. & Kadarmideen, H. (2008). Principal component and clustering analysis of functional traits in Swiss dairy cattle. *Turk. J. Vet. Anim. Sci* 32(3): 163-171.
- Khatami, a. (1984). Camel potential applications of scholarly perspectives. *Institute of Animal Husbandry*.
- Lee, S.-d. (2006). Characterisation of multiple interior noise metrics and translation of the voice of the customer. *International Journal of Vehicle Noise and Vibration* 2(4): 341-356.
- Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*. Lawrence Erlbaum Associates Publishers.
- Macciotta, N. P. P. & Gaspa, G. (2010). Use of Principal Component and Factor Analysis to reduce the number of independent variables in the prediction of Genomic Breeding Values. *Italian Journal of Animal Science* 8(2s): 105-107.
- Malau-Aduli, A., Aziz, M., Kojima, T., Niibayashi, T., Oshima, K. & Komatsu, M. (2004). Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle.
- Motulsky, H. (2009). Multicollinearity in multiple regression. <http://www.graphpad.com/articles/>.
- Ogah, D. (2011). Assessing size and conformation of the body of Nigerian indigenous turkey. *Slovak Journal of Animal Science* 44(1): 21-27.
- Pinto, L. & Dacker, I. (2006). De melo, CMR, Ledur, MC and Coutinho, LL 2006. Principal components analysis applied to performance and carcass traits in the chicken. *Anim. Res* 55: 419-425.
- RH, M. (1986). Classical and modern regression with applications. Duxbury Press, Boston, MA.
- Rosario, M., Silva, M., Coelho, A., Savino, V. & Dias, C. (2008). Canonical discriminant analysis applied to broiler chicken performance. *animal* 2(03): 419-424.
- Sangdon, L. (2006). multicollinearity in regression. *Journal of Agricultural Science* 5(9): 14.
- Smith, L. I. (2002). A tutorial on principal components analysis. *Cornell University, USA* 51: 52.
- Udeh, I. & Ogbu, C. (2011). Principal Component Analysis of Body Measurements In Three Strains of Broiler Chicken. *Science World Journal* 6(2): 11-14.
- Waring, M. E., Eaton, C. B., Lasater, T. M. & Lapane, K. L. (2010). Correlates of weight patterns during middle age characterized by functional principal components analysis. *Annals of epidemiology* 20(3): 201-209.
- Yakubu, A. & Idahor, K. O. (2009). Using factor scores in multiple linear regression model for predicting the carcass weight of broiler chickens using body measurements. *Revista Científica UDO Agricola* 9(4): 963-967.
- Yugfin, A. (2008). Variance Inflation and Orthogonalization in Regression. *Livestock Production Science* 52(6): 964-986.

■ ■ ■ ■ ■ ■ ■ ■ ■ ■ ■