

استفاده از روش تحلیل مولفه‌های اصلی برای پیش بینی صفت وزن دنبه در گوسفندان نژاد لری بختیاری

محمد رضا بختیاری زاده^{۱*}، محمد مرادی شهر بابک^۲، حسین مرادی شهر بابک^۳ و محمود وطن خواه^۴

۳،۲،۱ دانشجوی دکتری ژنتیک و اصلاح دام، استاد و استادیار گروه علوم دامی پردیس کشاورزی و منابع

طبیعی دانشگاه تهران، ۴، دانشیار دانشگاه آزاد اسلامی، واحد شهرکرد

(تاریخ دریافت: ۸۹/۳/۱۰ - تاریخ تصویب: ۹۰/۳/۲۵)

چکیده

به منظور بررسی ارتباط ۱۱ صفت وزن زنده، طول بدن حیوان، محیط دور بدن، قد حیوان، عرض وسط دنبه، عرض پایین دنبه، عرض بالای دنبه، طول دنبه، طول شکاف دنبه، عمق دنبه و محیط دور دنبه با صفت وزن دنبه از اطلاعات ۷۳۱ راس گوسفندان لری بختیاری و با هدف از بین بردن مشکل هم‌راستایی چندگانه، از روش‌های تحلیل مولفه‌های اصلی و حداقل مربعات معمولی استفاده گردید. وجود هم‌راستایی چندگانه با استفاده از عامل تورم واریانس (بیشتر از ۵ یا ۱۰) در بعضی از متغیرهای مستقل شناسایی شد. نتایج حاصل از این مطالعه نشان داد که مشکل هم‌راستایی چندگانه موجود در اطلاعات مربوط به ارتباط بین وزن دنبه گوسفند لری بختیاری با ۱۱ متغیر مستقل مربوط به این صفت با استفاده از روش تابعیت مولفه‌های اصلی قابل حل می‌باشد. همچنین از کل این ۱۱ متغیر صفات طول شکاف دنبه، عمق دنبه و محیط دور دنبه به ترتیب بالاترین ضریب را در برآورد وزن دنبه و به همین ترتیب صفات محیط دور بدن و طول دنبه به ترتیب کمترین ضریب را در برآورد وزن دنبه به خود اختصاص دادند.

واژه های کلیدی: هم‌راستایی چندگانه، تحلیل مولفه‌های اصلی، گوسفند لری بختیاری

مقدمه

کننده ممکن است با استفاده از روش کلاسیک حداقل مربعات برآورد شود. با توجه به وجود هم‌راستایی چندگانه در داده‌های مربوط به متغیرهای مستقل در مورد صفت وزن دنبه گوسفند، کاربرد روش حداقل مربعات معمولی (Ordinary Least Squares) به منظور برآورد ضرایب تابعیت متغیرهای مستقل منجر به ناپایداری و تغییرپذیری ضرایب تابعیت خواهد شد. زمانی که متغیرهای مستقل در برآورد میزان لاشه گوسفند، هم‌راستایی چندگانه نشان دهند، برآورد ضرایب تابعیت با استفاده از روش حداقل مربعات معمولی بزرگتر بوده، این ضرایب به میزان زیادی در علامت و مقدار، با تغییر کوچکی در متغیرهای مستقل

معمولا پژوهشگران در مورد موضوع مورد مطالعه خود، داده‌هایی را مربوط به چندین متغیر جمع آوری می‌کنند. داده‌های مربوط به این متغیرها را، داده‌های چند متغیره می‌نامند و از روش‌های آماری چند متغیره، برای تجزیه و تحلیل آنها استفاده می‌کنند. اگر متغیرهای مستقل با همدیگر همبسته باشند، وضعیتی تحت عنوان هم‌راستایی چندگانه رخ می‌دهد. در نتیجه نتایج بدست آمده از بعضی از روش‌های تجزیه‌ای مانند تابعیت ساده قابلیت اعتماد کمتری پیدا می‌کند. زمانی که متغیرهای مستقل هم‌راستایی (همبستگی) کمی نشان می‌دهند، ضرایب تابعیت برای تابع پیش‌بینی

تبدیل خطی متعامد است که داده را به دستگاه مختصات جدید می‌برد به طوری که بزرگترین واریانس داده بر روی اولین محور مختصات، دومین بزرگترین واریانس بر روی دومین محور مختصات قرار می‌گیرد و همین طور برای بقیه تحلیل مولفه‌های اصلی می‌تواند برای کاهش ابعاد داده مورد استفاده قرار گیرد، به این ترتیب مولفه‌هایی از مجموعه داده که بیشترین تاثیر را در واریانس دارند حفظ می‌کند. تابعیت مولفه‌های اصلی (Principal component regression) روشی است که به منظور برطرف کردن مشکل هم‌راستایی چندگانه و تولید برآوردهای ثابت و دقیق برای ضرایب تابعیت مورد استفاده قرار می‌گیرد. al. Fritts et. (۱۹۷۱)

برای اولین بار روش تابعیت مولفه‌های اصلی را برای برآورد ضرایب تابعیت در داده‌های با مشکل هم‌راستایی چندگانه ارائه کردند. برآورد کننده‌های پارامترها در تابع پیش‌بینی، بعد از انجام تابعیت مولفه‌های اصلی به عنوان برآورد کننده‌های مولفه‌های اصلی نامیده می‌شوند (Fritts et al., 1971).

این تبدیل متغیرهای متعامد جدید را بر اساس میزان اهمیت‌شان رتبه‌بندی می‌کند و سپس این روش بعضی از متغیرهایی که باعث کاهش واریانس می‌شوند را حذف می‌کند. بعد از حذف مولفه‌های اصلی با کمترین اهمیت، با استفاده از انجام تابعیت چند متغیره، متغیر وابسته در مقابل مجموعه کاهش یافته مولفه‌های اصلی، با استفاده از برآورد حداقل مربعات معمولی تجزیه می‌شود. با توجه به این که مولفه‌های اصلی متعامد هستند، دو به دو مستقل بوده و از این رو OLS مناسب می‌باشد.

زمانی که ضرایب تابعیت برای مجموعه کاهش یافته از متغیرهای متعامد محاسبه گردید، با یک تبدیل ریاضی به یک مجموعه جدید از ضرایب مطابق با مجموعه متغیرهای همبسته اولیه تبدیل می‌شوند. این ضرایب جدید برآورد کننده‌های مولفه‌های اصلی نامیده می‌شوند (Khalidari, 2008). هدف از انجام این تحقیق استفاده از روش تابعیت مولفه‌های اصلی به منظور بر طرف کردن مشکل هم‌راستایی چندگانه و برآورد ضرایب تابعیت چندگانه برای وزن دنبه در گوسفندان لری بختیاری بود.

یا وابسته، نوسان دارند و همچنین ضرایبی هستند با خطاهای استاندارد تورم یافته که در نهایت معنی‌دار نخواهند شد. مهم‌تر این که روش حداقل مربعات معمولی (OLS) میزان ضریب تعیین را برای مدل افزایش می‌دهد.

در نتیجه با وجود هم‌راستایی چندگانه در داده‌ها نمی‌توان از روش‌های تابعیت معمولی استفاده کرد و این مشکل در چهار خصوصیت مدل شامل مقدار، علامت، خطاهای استاندارد و ضریب تعیین که برای ما بسیار مهم می‌باشند، تاثیر منفی می‌گذارد (Jeeshim & Motulsky, 2002; Kucc, 2002).

وجود هم‌راستایی چندگانه فرض رتبه کامل بودن (Full rank) ماتریس $X'X$ را نقض می‌کند و در نتیجه انجام حداقل مربعات معمولی برای برآورد ضرایب تابعیت وجود ندارد. مدل تابعیت چند متغیره برابر با $y = X\beta$ بوده و برآورد کننده حداقل مربعات $\hat{\beta}$ برابر با $\hat{\beta} = (X'X)^{-1} X'Y$ می‌باشد. به منظور یافتن $\hat{\beta}$ باید ماتریس $(X'X)$ معکوس گردد. در اینجا $X'X$ مشابه ماتریس کواریانس یا همبستگی X می‌باشد که بستگی به این دارد که آیا ستون‌های ورودی نرمال یا استاندارد شده‌اند.

در صورتی که متغیرهای پیش‌بینی کننده همبستگی بالایی داشته باشند، خطای استاندارد ضرایب تابعیت جزئی برآورد شده زیاد بوده (به عبارتی واریانس برآوردها را افزایش می‌دهد) و تفسیر آنها نیز غیرممکن است، بنابراین تفسیر ساده ضرایب تابعیت جزئی بطوری که اثرات عمده را اندازه‌گیری کنند، تضمین شده نیست. هم‌راستایی چندگانه مشکلی است که در اکثر موارد وجود دارد و بر روی فرض معکوس ماتریس تاثیر می‌گذارد (Sangdon, 2006).

وجود هم‌راستایی چندگانه، که در ناپایداری ضرایب برآورد شده نمایان می‌گردد را می‌توان با استفاده از معیارهای آماری همچون ارزش تحمل (Tolerance value)، عامل تورم واریانس (Variance Inflation factor) و مقادیر ویژه (Eigenvalue) شناسایی کرد (Yu, 2008).

Pearson (۱۹۰۱) روش تحلیل مولفه‌های اصلی را ارائه داد. تحلیل مولفه‌های اصلی در تعریف ریاضی یک

مواد و روش‌ها

این تحقیق با استفاده از رکوردهای ۷۳۱ راس گوسفند لری بختیاری، که از اواسط خرداد ماه تا اواسط آذر ماه سال ۱۳۸۲ برای ذبح به کشتارگاه صنعتی جوفقان واقع در استان چهارمحال بختیاری آورده شدند، انجام گرفت. گوسفندان مورد آزمایش در این مطالعه به صورت تصادفی از بین گوسفندان ارائه شده برای فروش انتخاب گردیدند. اندازه‌های بدن و ابعاد ظاهری دنبه با استفاده از متر پارچه‌ای و با دقت یک سانتیمتر اندازه‌گیری و ثبت شدند. اندازه‌های بدن شامل طول بدن در ناحیه کمر، ارتفاع بدن و دور سینه بودند. ابعاد ظاهری دنبه شامل اندازه عرض دنبه در سه قسمت بالا، وسط و پایین، اندازه طول دنبه، اندازه طول شکاف دنبه، اندازه عمق دنبه و اندازه محیط بالای دنبه بودند.

پس از اندازه‌گیری‌ها، گوسفندان به روش مرسوم در کشتارگاه صنعتی ذبح شدند و بعد از خون‌گیری، پوست-کنی، تخلیه امعا و احشا از حفره بطنی، ارزیابی لاشه از نظر بهداشتی توسط دامپزشک صورت گرفت. لاشه گرم توزین و بعد از جدا کردن دنبه از بدن دوباره لاشه بدون دنبه توزین و ثبت گردید. پس از دسته‌بندی و وارد کردن داده‌ها به رایانه، برای تعیین بهترین معادله تابعیت تجزیه‌های تابعیت چندگانه و تحلیل مولفه‌های اصلی صورت گرفت. در تجزیه و تحلیل داده‌ها متغیرهای مستقل عبارت از وزن زنده، طول بدن حیوان، محیط دور بدن، قد حیوان (از جلوگاه تا زمین)، عرض وسط دنبه، عرض پایین دنبه، عرض بالای دنبه، طول دنبه، طول شکاف دنبه، عمق دنبه، محیط دور دنبه و متغیر وابسته نیز وزن دنبه بود.

برای انجام تابعیت چندگانه از نرم‌افزار SAS و رویه REG استفاده شد. مدل مورد استفاده به شرح زیر است: (Xها برابر با متغیرهای مستقل هستند که در بالا ذکر شد)

$$y_i = b_0 + b_{1i}(x_1) + \dots + b_{11i}(x_{11}) + e_i$$

در این مدل، y متغیر وابسته، x_i متغیرهای مستقل و b_i ضرایب تابعیت است. ابتدا همبستگی بین داده‌ها برآورد و بررسی شد. همچنین عامل تورم

واریانس (Variance Inflation Factor) (به منظور بررسی کنترل هم‌راستایی چندگانه) نیز مورد بررسی قرار گرفت. بدین منظور از رویه REG و PRINCOMP نرم‌افزار SAS استفاده قرار شد. به منظور انجام روش تابعیت مولفه‌های اصلی ابتدا داده‌ها طبق فرمول

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

متغیر مستقل استاندارد شده بوده و x_i و s_i به ترتیب میانگین و انحراف معیار هر متغیر مستقل مربوطه می‌باشند. بعد از استاندارد کردن داده‌ها، از رویه PRINCOMP نرم‌افزار SAS برای محاسبه مقادیر ویژه و بردارهای ویژه هر یک از مقادیر ویژه استفاده شد. پس از برآورد مقادیر و بردارهای ویژه مربوطه با استفاده از روش

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

معادله تابعیت جدید مورد استفاده قرار گرفت. معادله جدید 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 $prin$ ها در این معادله برابر با بردارهای ویژه مربوط به مقدار ویژه‌ی مربوطه و k برابر با تعداد مقدار ویژه بدست آمده از فرمول $\frac{\sum_{j=1}^k \lambda_j}{k} > 0.99$ می‌باشد. همچنین، y برابر با متغیر وابسته استاندارد شده و B ها ضرایب تابعیت مربوطه و B_0 نیز عرض از مبدا می‌باشند. پس از برآورد ضرایب تابعیت $prin$ ها با کاربرد معادله ماتریسی زیر ضرایب تابعیت متغیرهای مستقل استاندارد شده (برآورد کننده‌های مولفه‌های اصلی) بدست آمد.

$$\begin{bmatrix} b_{1.pc} \\ \vdots \\ b_{11.pc} \end{bmatrix} = \begin{bmatrix} prin_1 & \dots & prin_k \end{bmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_k \end{bmatrix}$$

در این معادله، b_{pc} ها برابر با ضرایب تابعیت متغیرهای مستقل استاندارد شده می‌باشند. برای بدست

ضرایب متغیرهای مستقل اولیه (قبل از استاندارد کردن) برابر است با:

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

که در این معادله $s.e.(b_{j.pc})$ خطای استاندارد برآورد کننده مولفه اصلی از ضریب مربوطه به j امین متغیر مستقل می‌باشد. یکی از معیارهای مهم برای مقایسه این دو روش، خطای استاندارد برآوردها می‌باشد، زیرا همانطور که بیان شد یکی از پیامدهای هم‌راستایی چندگانه تورم واریانس برآوردها می‌باشد، در نتیجه برآوردهای بدست آمده از روش تابعیت مولفه‌های اصلی، صحت بالاتر و خطای استاندارد کمتری خواهند داشت.

نتایج و بحث

خلاصه آماری صفات مورد استفاده، پارامترهای برآوردی و خطای استاندارد برای روش حداقل مربعات معمولی به ترتیب در جدول‌های ۱ و ۲ ارائه شده است. ضریب تعیین این برآوردها ۰/۸۳۸ بود. به منظور بررسی هم‌راستایی چند گانه ابتدا باید همبستگی‌های بین متغیرهای مستقل بررسی شود. در صورت وجود همبستگی بالا باید معیارهای معتبرتر دیگری را برای تایید وجود هم‌راستایی چند گانه به کار برد. در جدول ۳ همبستگی‌های بین متغیرهای مستقل ارائه شده است.

آوردن ضرایب تابعیت متغیرهای مستقل قبل از استاندارد کردن از فرمول

$$b_{j.pc}^* = \frac{b_{j.pc}}{s_j}, j = 1, 2, \dots, 11$$

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

$$b_{0.pc}^* = b_{0.pc} - \frac{b_{1.pc} \bar{x}_1}{s_1} - \frac{b_{2.pc} \bar{x}_2}{s_2} - \dots - \frac{b_{11.pc} \bar{x}_{11}}{s_{11}}$$

که $b_{0.pc}^*$ برابر با عرض از مبدا برای معادله تابعیت متغیرهای مستقل قبل از استاندارد کردن می‌باشد. برای محاسبه خطاهای استاندارد این برآوردها نیز از معادلات زیر استفاده شد.

$$\text{var}(b_{pc}^*) = \begin{bmatrix} \text{prin}_{11}^2 & \cdot & \cdot & \text{prin}_{1k}^2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \text{prin}_{11}^2 & \cdot & \cdot & \text{prin}_{1k}^2 \end{bmatrix} \begin{bmatrix} \text{var}(B_1) \\ \cdot \\ \cdot \\ \text{var}(B_{11}) \end{bmatrix}$$

در ماتریس بالا $\text{VAR}(B_i)$ برابر با توان ۲ انحراف معیارهای (واریانس) بدست آمده برای هر کدام از ضرایب تابعیت در مرحله اول می‌باشد. برآورد کننده‌های مولفه اصلی از ضرایب تابعیت در نهایت با معادله $b_{j.pc}^* = \frac{b_{j.pc}}{s_j}$ که s_j برابر با خطای استاندارد متغیر مستقل مربوطه می‌باشد، بدست می‌آید. در نتیجه، خطای استاندارد برآورد کننده‌های مولفه اصلی برای

جدول ۱- خلاصه آماری صفات

صفت	میانگین	انحراف معیار	صفت	میانگین	انحراف معیار
وزن زنده بدن (kg)	۴۴/۴۹	۱۴/۵۹	عرض پایین دنبه	۲۴/۶۷	۷/۱۷
طول بدن	۴۹/۹۸	۵/۲۸	طول دنبه	۲۶/۴۶	۵/۸۰
ارتفاع	۷۲/۴۲	۶/۹۶	طول شکاف دنبه	۱۴/۹۳	۳/۵۲
محیط دور بدن	۷۹/۹۰	۱۹/۰۵	عمق دنبه	۱۴/۰۸	۳/۱۹
عرض بالای دنبه	۱۵/۸۰	۴/۸۸	محیط دور دنبه	۵۲/۴۵	۱۱/۷۷
عرض وسط دنبه	۲۳/۷۹	۶/۹۴	وزن دنبه (kg)	۲/۶۸	۲/۲۵

* غیر از دو صفت مشخص شده، واحد اندازه‌گیری بقیه صفت سانتی متر می‌باشد.

جدول ۲- برآوردهای بدست آمده برای معادله تابعیت با استفاده از روش حداقل مربعات معمولی

خطای استاندارد	برآورد	صفت	خطای استاندارد	برآورد	صفت
۰/۰۲۴	۰/۰۸۵	عرض پایین دنبه	۰/۰۰۵	۰/۰۳۷	وزن زنده بدن (kg)
۰/۰۱۵	۰/۰۱۶	طول دنبه	۰/۰۱۰	-۰/۰۴۴	طول بدن
۰/۰۱۹	۰/۱۰۹	طول شکاف دنبه	۰/۰۰۹	-۰/۰۴۳	ارتفاع
۰/۰۲۰	۰/۰۹۷	عمق دنبه	۰/۰۰۱	-۰/۰۰۹	محیط دور بدن
۰/۰۰۶	۰/۰۶۸	محیط دور دنبه	۰/۰۱۲	-۰/۰۲۵	عرض بالای دنبه
-	-	-	۰/۰۲۵	۰/۰۰۷	عرض وسط دنبه

برای معکوس ماتریس $X'X$ به منظور برآورد پارامترها در روش حداقل مربعات معمولی به هم خورده و تورم واریانس ایجاد می‌شود پس با وجود خطای استاندارد بالا برای بعضی از برآوردها (مانند صفت عرض وسط دنبه) نیاز به یک روش بهتر برای برآوردهای دقیق‌تر می‌باشد. برای بررسی هم‌راستایی چند گانه عامل تورم واریانس نیز محاسبه شد (جدول ۴).

با مشاهده همبستگی‌های موجود بین متغیرها، همبستگی بالا بین اکثر این متغیرها مشهود است. در نتیجه با وجود این همبستگی‌ها باید هم‌راستایی چندگانه بررسی شود. برآورد ضریب تعیین بالای بدست آمده در این حالت می‌تواند ناشی از تورم واریانس حاصل از هم‌راستایی بوده باشد، زیرا در این حالت شرط نبود رابطه بین متغیرها

جدول ۳- همبستگی‌های بین متغیرهای مستقل

صفات	(۱)	(۲)	(۳)	(۴)	(۵)	(۶)	(۷)	(۸)	(۹)	(۱۰)	(۱۱)
وزن زنده بدن (۱)	۱	۰/۷۱	۰/۸۴	-۰/۰۸	۰/۴۷	۰/۶۷	۰/۶۸	۰/۷۲	۰/۶۶	۰/۶۵	۰/۶۶
طول بدن (۲)	-	۱	۰/۶۳	۰/۰۱	۰/۴	۰/۳۸	۰/۳۹	۰/۵۲	۰/۴۳	۰/۳۹	۰/۵۳
ارتفاع (۳)	-	-	۱	۰/۰۲	۰/۴۰	۰/۶۰	۰/۵۹	۰/۶۰	۰/۵۵	۰/۵۴	۰/۵۲
محیط دور بدن (۴)	-	-	-	۱	۰/۱۶	۰/۱۰	۰/۰۹	-۰/۰۳	-۰/۱۴	۰/۰۴	۰/۰۰
عرض بالای دنبه (۵)	-	-	-	-	۱	۰/۷۳	۰/۷۱	۰/۶۳	۰/۴۵	۰/۵۷	۰/۷۲
عرض وسط دنبه (۶)	-	-	-	-	-	۱	۰/۹۸	۰/۷۴	۰/۵۹	۰/۷۷	۰/۷۴
عرض پایین دنبه (۷)	-	-	-	-	-	-	۱	۰/۷۶	۰/۶۰	۰/۷۸	۰/۷۶
طول دنبه (۸)	-	-	-	-	-	-	-	۱	۰/۸۴	۰/۷۹	۰/۸۱
طول شکاف دنبه (۹)	-	-	-	-	-	-	-	-	۱	۰/۷۳	۰/۶۹
عمق دنبه (۱۰)	-	-	-	-	-	-	-	-	-	۱	۰/۷۷
محیط دور دنبه	-	-	-	-	-	-	-	-	-	-	۱

بحرانی ۵ تا ۱۰ می‌باشد) بوده و در این حالت هم‌راستایی چندگانه رخ خواهد داد. تورم واریانس نتیجه‌ای از وجود هم‌راستایی چندگانه است. در صورتی

با توجه به جدول ۴ مشاهده می‌شود که صفات عرض وسط دنبه، عرض پایین دنبه، وزن زنده و طول دنبه میزان تورم واریانس بیشتر از مقدار بحرانی (مقدار

که وجود این متغیرها در مدل ضروری باشد، باید از روش دیگری مانند تابعیت مولفه‌های اصلی استفاده کنیم. در این حالت قادر خواهیم بود برآوردهایی دقیق‌تر و با خطای استاندارد کمتر بدست آوریم. مرحله بعد محاسبه معادله تابعیت مورد نظر از طریق روش تابعیت مولفه‌های اصلی می‌باشد.

جدول ۴- مقادیر تورم واریانس برای متغیرهای مستقل

تورم واریانس	صفت	تورم واریانس	صفت
۲۶/۸۷	عرض پایین دنبه	۶/۳۳	وزن زنده بدن
۶/۲۳	طول دنبه	۲/۵۰	طول بدن
۴/۰۳	طول شکاف دنبه	۳/۶۷	ارتفاع
۳/۹۱	عمق دنبه	۱/۱۸	محیط دور بدن
۴/۷۴	محیط دور دنبه	۳/۰۲	عرض بالای دنبه
-	-	۲۷/۴۹	عرض وسط دنبه

ویژه اول همبسته نیست. و به همین ترتیب این دو خصوصیت برای مقادیر ویژه بعدی وجود دارد. در نتیجه با استفاده از تابعیت مولفه‌های اصلی داده را متعامد یا غیر همبسته کرده و سپس معادله تابعیت مورد نظر را بدست می‌آوریم. لازم به ذکر است که هر مقدار ویژه حاوی ۱۱ برآورد برای هر کدام از متغیرهای مستقل می‌باشد. و این برآوردها همان بردار ویژه‌ها می‌باشند. به عبارت دیگر در صورتی که تنها مقدار ویژه اول را انتخاب کنیم با ۱۱ برآورد بدست آمده برای هر کدام از متغیرها فقط ۰/۶۲ از واریانس توضیح داده می‌شود.

بدین منظور ابتدا باید مقادیر ویژه و بردارهای ویژه مربوطه را برآورد کرد. در جداول ۵ و ۶ به ترتیب مقادیر ویژه و بردارهای ویژه برآورد شده ارائه شده است. همانطور که در جدول ۵ مشاهده می‌شود، به ترتیب مقادیر ویژه از بالا به پایین کاهش می‌یابند. همچنین با کاهش ضرایب مقادیر ویژه میزان سهم آنها در توضیح واریانس نیز کاهش می‌یابد. بدین ترتیب مقدار ویژه اول بالاترین سهم از کل واریانس (۰/۶۲) و مقدار ویژه آخر نیز کمترین مقدار از کل واریانس (۰/۰۱) را توضیح می‌دهند. مقدار ویژه دوم نیز بخشی از واریانس کل را که مقدار ویژه اول توضیح نداده، در بر می‌گیرد و با مقدار

جدول ۵- مقادیر ویژه برای متغیرهای مورد استفاده

ردیف	مقادیر ویژه	واریانس*	ردیف	مقادیر ویژه	واریانس
۱	۶/۸۱	۰/۶۲	۷	۰/۲۱	۰/۰۲
۲	۱/۲۲	۰/۱۱	۸	۰/۱۶	۰/۰۱
۳	۰/۹۸	۰/۰۹	۹	۰/۱۲	۰/۰۱
۴	۰/۵۶	۰/۰۵	۱۰	۰/۱۱	۰/۰۱
۵	۰/۵۳	۰/۰۵	۱۱	۰/۰۲	۰/۰۱
۶	۰/۲۷	۰/۰۲	-	-	-

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

عنوان مثال، مقدار ویژه اول برابر با ۶/۸۱ می‌باشد، به عبارت دیگر این مقدار ویژه ۶/۸۱ واحد از ۱۱ واحد واریانس را توضیح می‌دهد که درصد آن برابر ۰/۶۲ می‌باشد 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 تجزیه مورد بررسی قرار می‌گیرند و در این مطالعه برابر با ۱۱ (عدد ۱۱ از جمع واریانس ۱۱ متغیر مستقل با واریانس ۱ بدست آمده است) می‌باشد. در نتیجه به

جدول ۶- بردارهای ویژه برای متغیرهای مورد استفاده

صفات	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10
وزن زنده بدن	۰/۳۲۶	-۰/۲۹۰	۰/۲۵۵	-۰/۰۱۶	-۰/۲۱۵	-۰/۰۰۳	-۰/۰۳۲	-۰/۰۷۵	۰/۱۵۹	-۰/۸۱۴
طول بدن	۰/۲۴۳	-۰/۳۰۶	۰/۵۳۷	-۰/۲۲۱	۰/۴۴۰	-۰/۳۶۲	۰/۲۸۴	۰/۲۰۲	۰/۰۴۴	۰/۲۴۳
ارتفاع	۰/۲۸۹	-۰/۲۳۹	۰/۴۰۵	۰/۰۴۶	-۰/۴۳۹	۰/۳۷۷	-۰/۳۷۵	-۰/۰۶۹	-۰/۱۴۸	۰/۴۳۴
محیط دور بدن	۰/۰۰۹	۰/۶۹۱	۰/۵۵۵	۰/۴۳۳	۰/۰۸۹	۰/۰۲۳	۰/۰۳۴	-۰/۰۷۶	۰/۰۵۳	-۰/۰۸۴
عرض بالای دنبه	۰/۲۸۴	۰/۳۴۶	-۰/۰۵۱	-۰/۵۲۶	۰/۳۰۶	۰/۴۷۲	-۰/۲۱۸	۰/۳۶۷	۰/۰۲۸	-۰/۱۱۶
عرض وسط دنبه	۰/۳۳۹	۰/۲۴۵	-۰/۱۱۵	-۰/۱۸۳	-۰/۳۶۱	-۰/۱۰۶	۰/۳۲۵	۰/۰۰۸	۰/۱۲۹	۰/۱۰۳
عرض پایین دنبه	۰/۳۴۲	۰/۲۲۸	-۰/۱۲۷	-۰/۱۶۰	-۰/۳۴۱	-۰/۱۷۵	۰/۳۵۱	-۰/۰۹۶	۰/۰۵۹	۰/۱۳۶
طول دنبه	۰/۳۴۶	-۰/۰۴۵	-۰/۱۴۷	۰/۲۳۶	۰/۲۱۴	۰/۱۶۹	۰/۲۸۶	-۰/۰۹۹	-۰/۷۶۸	-۰/۱۱۰
طول شکاف دنبه	۰/۳۰۵	-۰/۲۰۱	-۰/۲۳۵	۰/۵۱۹	۰/۲۱۶	۰/۳۶۸	۰/۲۱۴	۰/۱۳۴	۰/۵۲۴	۰/۱۳۴
عمق دنبه	۰/۳۲۹	۰/۰۸۲	-۰/۲۰۴	۰/۳۰۰	-۰/۰۴۹	-۰/۵۱۴	-۰/۴۸۷	۰/۴۸۸	-۰/۰۹۵	۰/۰۰۴
محیط دور دنبه	۰/۳۳۸	۰/۰۶۶	-۰/۱۴۴	-۰/۰۹۱	۰/۳۴۹	-۰/۱۷۸	-۰/۳۶۴	-۰/۷۲۹	۰/۱۵۲	۰/۰۸۶

* با توجه به این که ۱۰ بردار ویژه اول ۰/۹۹ واریانس را توضیح می‌دهند و از این ۱۰ بردار ویژه استفاده شد، بردار ویژه ۱۱ در این جدول نشان داده نشده است.

استاندارد آنها نشان داده شده است. در مرحله بعد باید این برآوردها را به مقادیر اولیه تبدیل کرد. بدین منظور 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 گردد. در جدول ۸ این برآوردها به همراه خطاهای استاندارد مربوطه ارائه شده است.

با توجه به برآوردهای صورت گرفته در این مطالعه بر اساس روش تابعیت مولفه‌های اصلی، صفات طول شکاف دنبه، عمق دنبه و محیط دور دنبه به ترتیب بالاترین ضریب را در برآورد وزن دنبه به خود اختصاص دادند و به همین ترتیب صفات محیط دور بدن و طول دنبه به ترتیب کمترین ضریب را در برآورد وزن دنبه در

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 و دوم بیشترین واریانس را توضیح داده و بقیه مقادیر ویژه نقش کمتری را دارند که این دلیلی دیگر بر وجود همراستایی چندگانه می‌باشد. به منظور انتخاب تعداد مقادیر ویژه مورد نیاز برای انجام تابعیت مولفه‌های اصلی از فرمول $\frac{\sum_{j=1}^r \lambda_j}{k} > 0.99$ استفاده شد و در نتیجه ۱۰ بردار ویژه (پربین) از ۱۱ بردار ویژه موجود مورد استفاده قرار گرفت. در جدول ۷ این برآوردها به همراه خطاهای

برآوردها می‌باشد (لازم به ذکر است که مقایسه خطاهای استاندارد در این دو روش در شرایط یکسانی صورت گرفت. زیرا برآوردهای بدست آمده در روش تحلیل مولفه‌های اصلی به حالت قبل از استاندارد کردن تبدیل شد) بدین منظور با مشاهده جداول ۲ و ۸ می‌توان به این موضوع پی برد.

نژاد لری بختیاری بدست آوردند. در مورد روش حداقل مربعات معمولی این مورد برای صفات با بیشترین اهمیت یکسان بود اما در مورد صفات با کمترین اهمیت تفاوت داشت، بدین ترتیب که در این روش صفات عرض وسط دنبه و محیط دور بدن به ترتیب کمترین اهمیت را در برآورد وزن دنبه نشان دادند. برای مقایسه این دو روش بهترین راه مقایسه خطاهای استانداردهای

جدول ۷- برآوردهای بدست آمده برای معادله تابعیت با استفاده از تابعیت مولفه‌های اصلی

ردیف	خطای استاندارد	برآورد	ردیف	خطای استاندارد	برآورد
Prin 1	۰/۰۰۶	۰/۳۳۲	Prin 6	۰/۰۳۰	-۰/۱۵۵
Prin 2	۰/۰۱۴	-۰/۰۱۴	Prin 7	۰/۰۳۳	-۰/۰۲۴
Prin 3	۰/۰۱۵	-۰/۲۵۱	Prin 8	۰/۰۳۷	-۰/۲۴۳
Prin 4	۰/۰۲۰	۰/۰۶۵	Prin 9	۰/۰۴۳	۰/۱۶۵
Prin 5	۰/۰۲۰	۰/۰۰۱	Prin 10	۰/۰۴۴	-۰/۱۸۰

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

خطای استاندارد	برآورد	صفت	خطای استاندارد	برآورد	صفت
۰/۰۰۲	۰/۰۲۲	عرض پایین	۰/۰۰۳	۰/۰۱۶	وزن زنده بدن (kg)
۰/۰۰۶	۰/۰۰۸	طول دنبه	۰/۰۰۴	-۰/۰۱۹	طول بدن
۰/۰۰۸	۰/۰۴۷	طول شکاف	۰/۰۰۴	-۰/۰۱۹	ارتفاع
۰/۰۰۹	۰/۰۴۲	عمق دنبه	۰/۰۰۰۸	-۰/۰۰۴	محیط دور بدن
۰/۰۰۳	۰/۰۳۱	محیط دور دنبه	۰/۰۰۵	-۰/۰۱۳	عرض بالای دنبه
-	-	-	۰/۰۰۲	۰/۰۱۹	عرض وسط دنبه

ایجاد یک معادله تابعیت مناسب نمی‌باشند و برآوردهای صورت گرفته توسط روش تابعیت مولفه‌ها اصلی که بر مبنای ایجاد داده‌ها متعامد و سپس برآورد پارامترها عمل می‌کند، مناسب‌تر می‌باشد. (Malau et al. ۲۰۰۳) طی مطالعه‌ای برآورد معادله تابعیت را با روش‌های تابعیت ریج و مولفه‌های اصلی برای صفت وزن بدن در گاوهای سیاه ژاپنی با استفاده از ۱۳ متغیر مستقل بررسی کردند. نتیجه گرفتند که استفاده از تجزیه مولفه‌های اصلی برای برآورد ضرایب تابعیت پایدار و به منظور غلبه بر مشکل هم‌راستایی چندگانه مناسب می‌باشد.

با توجه به وجود هم‌راستایی چندگانه در داده‌ها، برآوردهای بدست آمده از روش حداقل مربعات معمولی

یکی از خصوصیات روش تابعیت مولفه‌های اصلی برآوردهای با خطای استاندارد کمتر نسبت به روش حداقل مربعات می‌باشد و این موضوع در این مطالعه نیز تایید شد. یکی از علت‌های مهم در این رابطه، از بین بردن همبستگی بین متغیرها و تبدیل آنها به متغیرهای متعامد می‌باشد. خطاهای استاندارد روش تابعیت مولفه‌های اصلی نسبت به روش حداقل مربعات کمتر برآورد شد. همچنین موضوع دیگر در این مورد این است که با توجه به این که در روش حداقل مربعات معمولی متغیرها نامتعامد می‌باشند، در نتیجه برآوردها دقت لازم را ندارند و اریب هستند و ممکن است با اضافه شدن داده‌های دیگر به مجموعه داده‌های فعلی، این برآوردها تغییر کنند و به این ترتیب، این برآوردها برای

هم‌راستایی چندگانه موجود در اطلاعات مربوط به ارتباط بین وزن دنبه گوسفند لری بختیاری با ۱۱ متغیر مستقل مربوط به این صفت با استفاده از روش تابعیت مولفه‌های اصلی قابل حل می‌باشد. همچنین از کل این ۱۱ متغیر صفات طول شکاف دنبه، عمق دنبه و محیط دور دنبه به ترتیب بالاترین ضریب را در برآورد وزن دنبه و به همین ترتیب صفات محیط دور بدن و طول دنبه به ترتیب کمترین ضریب را در برآورد وزن دنبه در نژاد لری بختیاری به خود اختصاص دادند.

ممکن است با تغییر داده‌ها (کم یا زیاد شدن داده‌ها) تغییر کند و در نتیجه معتبر نمی‌باشد. در صورت وجود هم‌راستایی چندگانه در داده‌ها بهتر است از روش تابعیت مولفه‌های اصلی برای برآورد معادله تابعیت نسبت به روش حداقل مربعات معمولی استفاده کرد. در صورت استفاده از روش تابعیت مولفه‌های اصلی، با توجه به قدرت این روش در مورد تبدیل متغیرهای نامتعامل به متغیرهای تعامل، خطاهای استاندارد برآوردها نسبت به روش حداقل مربعات معمولی، کمتر بدست می‌آید. در نهایت نتایج حاصل از این مطالعه نشان داد که مشکل

REFERENCES

1. Desta Fekedulegn B., Colbert J. J. Hicks R. R. Michael J. Schuckers E. (2002). *Coping with multicollinearity: an example on application of principal components regression in dendroecology*. USDA Forest Service, from <http://www.fs.fed.us/ne>.
2. 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: 845-864.
3. Jeeshim and Kucc. (2002). *Multicollinearity in Regression Models*. Retrieved January 25, 2008, from <http://www.masil.org/documents/multicollinearity.pdf>.
4. Khaldari M. (2008). *Sheep and goat husbandry* (3rd ed.). Tehran: Jahad Publication Organization. (In Farsi)
5. Malau-Aduli A. E. O., Aziz M. A. Kojima T. Niibayashi T. Oshima K. & Komatsu M. (2003). Fixing collinearity instability using principal component and ridge regression analyses in the relationship between body measurements and body weight in Japanese Black cattle. Retrieved January 5, 2008, from <http://www.eprints.utas.edu.au/588/2/JAVA2004.pdf>
6. Motulsky H. (2002). Multicollinearity in multiple regression. Retrieved January 29, 2008, from <http://www.graphpad.com/articles/Multicollinearity.htm>.
7. Pearson K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559-572.
8. Sangdon L. (2006). Multicollinearity in regression. Retrieved January 9, 2008, from <http://newsgroups.derkeiler.com/pdf/Archive/Sci/sci.stat.consult/2006-03/msg00168.pdf>.
9. SAS (2008). Statistical Analysis System, SAS Institute Inc., Cary, North Carolina, USA.
10. Yu A. (2008). Multi-collinearity, Variance inflation and orthogonalization in regression. Retrieved January 19, 2008, from http://www.creative-wisdom.com/computer/sas/collinear_VIF.html.

