Investigation of Live Weights at Different Ages by Cluster Analysis in Konya Merino Sheep

Birol Dağ
Department of Animal Science,
Faculty of Agriculture,
Selcuk University, 42075, Konya, Turkey
bdag@selcuk.edu.tr

İsmail Keskin
Department of Animal Science,
Faculty of Agriculture,
Selcuk University, 42075, Konya, Turkey
ikeskin@selcuk.edu.tr

Abstract: The aim of this study was to investigate the live weights at birth (BW), weaning (WW), sixth (SW), twelfth (TW) and eighteenth (EW) months of age by cluster analysis in Konya Merino sheep. k-means clustering methods was used for the cluster analysis. Clusters were obtained differently according to sex and birth types.

Effects of dam age on BW, WW ($P<0.01$) and SW ($P<0.05$) were found to be significant and its effects on TW and EW were not significant in single born females. Dam age did not affect BW, WW, SW, TW and EW in twin born females and single born males. In twin born males, dam age affected only SW ($P<0.05$).

By the statistically analyses different cluster numbers were determined as 6, 5 and 4 for single born females, twin born females, single and twin born males respectively. The differences between the clusters in respect of live weights at the same ages were statistically significant ($P<0.01$).

At the end of the study, according to live weights at eighteenth month of age, the third cluster for twin males and the first cluster for single females gave the highest values.

Keywords: Cluster Analysis, k-means, Konya Merino

Introduction

The relationships among birth weight which is an indicator of prenatal growth and weaning weight and also its relation to lamb survival rate up to weaning period and body weights at later ages are well known. Birth weight is one of the features that can be detected early. Lamb body weight at the onset of fattening period generally refers to weaning weight.

The cluster analysis is one of the multivariate statistical analysis and the main goal of it is to dividing a data set in hand into two or more clusters by taking into account a certain similarity measure. This splits are desired as homogenous within cluster and heterogenic between clusters (Hair et al., 1998).

This study was conducted to examine the body weights at various periods of Konya Merino sheep with cluster analysis.

Materials and Methods

Data from 162 Konya Merino lambs born in 2002 (57 males and 105 females) were used. The sheep were maintained at the Bahri Dağdaş Agricultural Research Institute farm in the Konya Province in central Turkey (37^{o}, $51^{'}$ N and 32^{o}, $48^{'}$ E). This province has approximately 1.6 million head of sheep, which represents 6% of the sheep population of Turkey. The province's average annual rainfall is between 250 - 400 mm; the mean temperature 11.5 °C; and the average elevation 1016 m.

The flock consisted of 400 ewes and 25 rams. Animals were maintained under semi-intensive conditions. Age at first lambing was approximately 24 months. Ewes lambed between 1 January and 15 February. The lambs were weaned at 75 days of age. Ewes were grazed from April to December and kept indoors throughout the winter. Live weight data were recorded monthly from birth to 16 months of age. Lambs were weaned at approximately 75 days of age and body weights at weaning, 6^{th}, 12^{th} and 18^{th} months of age were measured with an accuracy of 100 g. After 16 months the animals were kept for breeding purposes and no further weights were recorded.

Least square analysis were used to determine the dam effect on BW, WW, SW, TW and EW. It has been assumed that there is no interaction between the factors examined and the influences was determined by using the following statistical model in Harvey's (1987) package program.

$$Y_{ij} = \mu + a_i + e_{ij}$$

In the model;

Y_{ij} : observation for each trait,

μ : mean,

a_i : effect of dam age,

e_{ij} : random residual effect.

In clustering methods, there are two basic methods called hierarchical and nonhierarchical clustering methods are used when the units or variables are appropriately grouped according to clustering approach (Özdamar, 1999).

Different approaches are applied to combine the units each other in hierarchical clustering methods. These methods are commonly used and known are follows. 1. Single Linkage Method, 2. Average Linkage Method, 3. Complete Linkage Method, 4. McQuitty Linkage Method, 5. Centroid Linkage Method, 6. Median Linkage Method, 7. Ward Linkage Method.

In respect of having stronger theoretical basis than hierarchical clustering method and having prior information about cluster number provide preferring the nonhierarchical clustering methods to the hierarchical methods (Doğan, 2002).

The most used methods among nonhierarchical methods are maximum likelihood, k-means. K-means method were used in this study. Cluster number were determined by using the following equation.

$$k = \sqrt{\frac{n}{2}}$$

Where; n: number of units divided into clusters and k: cluster number (Tatlıdil, 1996).

When dividing variables into clusters, Mahalanobis distance is used as distance measure (Doğan, 2002). Mahalanobis distance is a generalized form of Euclid distance. Euclid distance between the units of a $n*p$ dimensional data matrix is:

$$d(i,j) = \sum_{k=1}^{P} (X_{ik} - X_{jk})^2 \qquad i = 1,2,.....n,$$

$$j = 1,2,.....n,$$

Where; n: number of units, p: number of variables, X_{ik} and Y_{jk} represent the values of k^{th} features of i^{th} and j^{th} units. Mahalanobis distance,

$$d^2(i,j) = (\overline{X}_i + \overline{X}_j)^t S^{-1} (\overline{X}_i - \overline{X}_j)$$

$\overline{X}_i$: observation vector of i^{th} units,

$\overline{X}_j$: observation vector of j^{th} units,

t: vector transpose, S^{-1} shows the inverse of similarity matrix.

In order to understand whether or not the used method constitute different clusters from each other, analysis of variance were made separately for all variables. Minitab 10.0 statistical software program were used for both clustering and analysis of variance.

Results and Discussion

At the end of the study, dam age on BW, WW ($P<0.01$) and SW ($P<0.05$) were found to be significant and its effect on TW and EW were not significant in single born females. Dam age did not affect BW, WW, SW, TW and EW in twin born females and single born males. In twin born males, dam age affected only SW ($P<0.05$).

This study in which the Konya Merino lambs divided into clusters by k-means method, cluster analysis were made by considering both sex and birth type which can affect clusters. Clustering results obtained by k-means method for single born female Konya Merino lambs were given in Table 1 and for twin born female lambs in Table 2. It is possible that both single and twin born female lambs can be selected for the investigated traits by taking into account the clusters.

If body weight is important for breeding, the first cluster for single born female lambs must be selected in respect of BW, WW, SW, TW and EW. But for twin born females in terms of BW and WW; SW; TW and EW, the third, fifth, first and fifth clusters must be chosen respectively.

Clustering results obtained by k-means method for single born male Konya Merino were given in Table 3 and for twin born male lambs in Table 4.

In terms of BW, WW, SW, TW and EW the first cluster for single born males must be selected. If both body weight and litter size are taken into consideration, the third cluster must be must be chosen from twin born males in terms of BW, WW, SW, TW and EW.

In order to understand whether the used method really divided into different clusters from each other, analysis of variance were made separately for all traits. The results of variance analysis showed that the difference between the formed clusters for both sex and birth types were statistically ($P<0.01$).

Cluster	BW	WW	SW	TW	EW
1	5.20±0.566	23.60±0.994	38.9±1.984	56.40±0.231	63.85±2.763
2	4.38±0.476	22.10±2.268	34.36±2.161	48.92±2.517	56.52±2.971
3	4.77±0.320	17.81±1.818	30.23±1.098	44.18±1.632	52.76±2.188
4	3.78±0.085	15.84±1.688	25.46±2.144	31.54±2.952	38.52±3.065
5	4.24±0.349	16.830±1.787	25.88±2.821	39.98±2.389	46.92±2.811
6	4.57±0.890	20.81±1.834	30.86±1.710	41.39±2.263	46.87±1.442

Table 1. Means and their Standard Deviations (Female/Single)

Cluster	BW	WW	SW	TW	EW
1	3.46±0.424	17.47±1.260	28.72±2.692	45.06±2.046	53.52±2.448
2	4.10±0.447	16.99±2.093	24.17±2.582	39.31±2.357	45.01±2.909
3	4.39±0.383	21.23±2.016	31.30±1.784	43.94±2.209	54.84±3.504
4	2.80±0.361	16.17±1.288	25.87±3.249	37.87±1.372	43.33±1.361
5	3.96±0.422	21.97±1.798	36.44±2.468	50.98±2.289	58.03±1.628

Table 2. Cluster Means and their Standard Deviations (Female/Twin)

Cluster	BW	WW	SW	TW	EW
1	5.45±0.296	26.67±1.514	43.20±4.025	58.42±4.307	70.18±2.332
2	4.54±0.197	23.02±3.441	36.65±1.925	56.08±2.590	71.39±4.592
3	2.70±0.021	14.00±0.145	21.90±0.188	32.00±0.204	47.30±0.237
4	4.78±0.612	20.45±4.669	32.07±4.092	48.92±3.129	60.70±5.012

Table 3. Cluster Means and their Standard Deviations (Male/Single)

Cluster	BW	WW	SW	TW	EW
1	3.63±0.231	15.20±0.222	25.60±3.298	43.93±5.525	54.33±3.791
2	3.96±0.493	19.81±1.869	32.45±2.146	55.00±3.960	68.74±3.003
3	4.67±0.577	25.10±2.972	39.33±1.502	65.33±1.865	77.50±1.412
4	3.87±0.363	18.05±1.777	30.79±2.215	45.40±2.985	57.90±3.881

Table 4. Cluster Means and their Standard Deviations (Male/Twin)

Conclusion

The first cluster not only has the highest value for BW, but also has the highest values WW, SW, TW and EW. According to this result, it can be said that the selection which will be made for BW can give reliable outcomes for the further age body weights. If the third cluster for twin born males and the fifth cluster for twin born females are selected in respect of BW, also the highest values for WW, SW, TW and EW could have been chosen.

In conclusion, it can be said that because of the advantages in terms of time, labor and cost, using the cluster analysis for the selection purposes would be appropriate than the other methods.

References

Doğan, İ. (2002). Kümeleme Analizi İle Seleksiyon (Selection by Cluster Analysis). *Turk J Vet Anim Sci*, 26: 47-53.

Hair, F.H., Andersen, R.E., Tahtam, R.L. & Black, W.C. (1998). Multivariate Data Analysis. Prentice Hall, New Jersey.

Harvey, W.R. (1987). User's Guide for LSMLMW PC-1 Version Mixed Model Least Squares and Maximum Likelihood Computer Program. Ohio State Univ., Columbus, Mimeo.

Minitab, (1995). Minitab Reference Manual, Release 10 Xtra. Minitab Inc. State Coll., PA 16801, USA.

Özdamar, K. (1999). Paket Programlar ile İstatistiksel Veri Analizi (Çok Değişkenli Analizler) II (Multivariate Analysis) . Kaan Kitabevi (2. Baskı), Eskişehir.

Tatlıdil, H. (1996). Uygulamalı Çok Değişkenli İstatiksel Analiz (Applied Multivariate Analysis), Ankara.