



Clustering of Counties of Samsun According to the Presence of Cattle Breeds

Burcu KURNAZ^{1*}, Hasan ÖNDER¹

¹ Ondokuz Mayıs University, Faculty of Agriculture, Department of Animal Science, 55139, Samsun, Türkiye.

Keywords:	ABSTRACT
Cattle breeds 3-D Clustering Principal Component Analysis Accuracy	In Samsun which is one of the leading provinces in cattle population, there is a great diversity in terms of cattle breeds. In this study, the cluster analysis method was used to examine the applicability of similar breeding policies for all districts when the statistical distribution of cattle in Samsun according to breeds was examined. The dimension reduction using PCA method was used to eliminate the variables that have no effect on the data set and to prevent the multicollinearity problem. As a result of the analysis, the explanatory power of the cluster analysis increased from 51.7% to 64.1% by using the dimension reduction. Different breeding policies should be developed for Bafra, Vezirköprü, Çarsamba and Tekkeköy districts compared to other districts, according to the results obtained by looking at the clustering graph formed when the ineffective variables are removed from the data set with the dimension reduction method. Policies for dairy cattle breeding can be proposed in these districts.

Samsun İlçelerinin Büyükbaş İrk Varlığına Göre Kümelmesi

Anahtar Kelimeler:	ÖZET
Sığır ırkları 3-D Kümeleme Temel Bileşenler Analizi Etkinlik	Büyükbaş hayvan popülasyonunda önde gelen illerden biri olan Samsun'da sığır ırkları açısından büyük bir çeşitlilik bulunmaktadır. Bu çalışmada Samsun ilindeki büyükbaş hayvanların ırklara göre istatistiksel dağılımı incelendiğinde benzer yetiştirme politikalarının tüm ilçeler için uygulanabilirliğinin incelenmesi amacıyla kümeleme analizi yöntemi kullanılmıştır. Veri seti üzerinde etkisi olmayan değişkenlerin ortadan kaldırılması ve çoklu bağlantı sorununun önlenmesi amacıyla PCA yöntemi kullanılarak boyut indirgeme işlemi uygulanmıştır. Analiz sonucunda boyut indirgeme kullanılarak kümeleme analizinin açıklama gücü %51,7'den %64,1'e çıkmıştır. Boyut indirgeme yöntemiyle etkisiz değişkenlerin veri setinden çıkarılmasıyla oluşan kümeleme grafiğine bakılarak elde edilen sonuçlara göre Bafra, Vezirköprü, Çarsamba ve Tekkeköy ilçeleri için diğer ilçelere göre farklı yetiştirme politikaları geliştirilmelidir. Bu ilçelerde süt sığırcılığına yönelik politikalar önerilebilir.

*Corresponding Author/Sorumlu Yazar: Burcu KURNAZ (burcukurnazz@hotmail.com)

Received: 23/02/2024

Accepted: 13/06/2024

Published: 30/06/2024

How to cite:

Kurnaz, B., Önder, H., (2024). Clustering of Counties of Samsun According to the Presence of Cattle Breeds. *Agro-Science Journal of Iğdır University*, 2(1), 1-6.



This work is licensed (CC BY-NC-ND 4.0) under Creative Commons Attribution 4.0 International License

INTRODUCTION

The livestock sector has a strategic importance in the adequate and balanced nutrition of the rapidly increasing population in Turkey, as in the world, in using it as an industrial raw material and in realizing rural development (Aral, 1996; Yılmaz and Köknaroglu, 2007). The aim of livestock policies and supports is to increase the production required for a balanced diet in terms of animal protein, and to realize a higher quality and economical production (Demir, 2012).

The growing food crisis all over the world underlines the need to take important strategic steps in the field of agriculture and animal husbandry. It is important for countries to develop productive policies in line with their geography and food source. Since cattle breeding is used both in food and industrial areas in our country, where animal husbandry is also very common in agriculture, the implementation of priority policies in order to ensure its continuity will continue the development in these areas.

It is necessary to repeat some studies in order to identify and solve the problems in livestock enterprises, to monitor the changes in the sector and to make realistic plans for the future (Ören and Bakır, 2020).

Statistical methods are used for development policies in many fields such as agriculture. Appropriate statistical analyzes are applied according to the variable structures obtained as a result of the experiments. Cluster Analysis, one of these analysis methods, is a multivariate statistical method that aims to divide a set of observations into a limited number of groups or clusters. Separation is made so that while observations in the same group are similar to each other, observations in different groups are different from each other (Neil, 2002).

In this study, it is aimed to classify the provinces of Samsun city according to cattle breeds in Turkey by clustering analysis on the applicability of similar breeding policies for districts according to cattle breed existence.

MATERIALS AND METHODS

In this study, a total of 64 cattle breeds from 17 counties in Samsun (Ondokuz Mayıs, Alaçam, Asarcık, Atakum, Ayvacık, Bafra, Canik, Çarşamba, Havza, İlkadım, Kavak, Ladik, Salpazarı, Tekkeköy, Terme, Vezirköprü, Yakakent) 392,289 cattle were used for analysis.

The cattle breeds used in the study; Aberdeen Angus, Aberdeen Angus Crossbreed, Angler, Angler Crossbreed, Aubrac, Avrupa Kırmızısı, Bazadaise Crossbreed, Belçika Mavisi, Belçika Mavisi Crossbreed, Blonde D'aquitaine, Blonde D'aquitaine Crossbreed, Boz Irk, Boz Irk Crossbreed, Brangus, Brangus Crossbreed, Brown Swiss, Brown Swiss Crossbreed, Charolais, Charolais Crossbreed, Chianina, Clavel Aleman, Clavel Aleman Crossbreed, Dogu Anadolu Kırmızısı, Dogu Anadolu Kırmızısı Crossbreed, Danimarka Kırmızısı Crossbreed, Eston Kırmızısı, Güney Anadolu Kırmızısı, Güney Anadolu Kırmızısı Crossbreed, Hereford, Hereford Crossbreed, Holstein KA, Holstein SA, Holstein KA M, Holstein SA, Holstein SA M, İsveç Kırmızısı, İsveç Kırmızısı Crossbreed, Jersey, Jersey Crossbreed, Leton Kırmızısı Crossbreed, Limousin, Limousin Crossbreed, Marchigiana Crossbreed, Montbeliarde, Montbeliarde Crossbreed, Normande, Normande Crossbreed, Norveç Kırmızısı, Norveç Kırmızısı Crossbreed, Piedmentosa, Piedmentosa Melezi, Red Angus, Red Angus Crossbreed, Salers, Salers Crossbreed, Simental, Simental Crossbreed, Yerli, Yerli Güney Sarısı, Yerli Güney Crossbreed, Yerli Kara, Yerli Kara Crossbreed, Zavot, Zavot Crossbreed, Zebu variables consisting breeds were used.

Cluster analysis, which is accepted as one of the multivariate statistical methods, is applied to group many complex data and compare the resulting groups, and it is often preferred because it is easy and the results can be clearly understood (Tekin, 2015). Cluster analysis is a technique that allows

data to be collected in discrete clusters in terms of their similarity to each other according to units or variables (Çakmak, 1999).

The main goal of cluster analysis is to classify and make sense of a group of data whose origin is unknown. Therefore, cluster analysis is used to classify units or objects according to their basic properties (Abonyi and Feil, 2007). In short, it can be stated that the general purpose of clustering analysis is to separate the similar from the different (Everitt et al., 2001).

In cluster analysis, the $N \times p$ dimensional data matrix, in which p measurements are made in each of N observations, can be shown as follows.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix}$$

Here, x_{ij} represents the value taken by the j^{th} variable for the i^{th} individual or object (Çakmak, 1999).

The six steps of Cluster Analysis;

Obtain the data matrix,

Standardize the data matrix,

Compute the resemblance matrix,

Execute the clustering method,

Rearrange the data and resemblance matrices,

Compute the cophenetic correlation coefficient. All else is supplementary detail (Romesburg, 2004).

Principal component analysis (PCA), which is a transformation technique that enables the dimensions of the data set, which includes a large number of interrelated variables, to be reduced to a smaller size by preserving the existing changes in the data as much as possible, facilitates various evaluations (Çilli, 2007). The analysis aims to determine the best transformation that can express the available data with fewer variables. The variables obtained after the transformation are called the principal components of the initial variables. The first principal component has the largest variance value and the other principal components are ordered in descending order of variance values. (Alpar, 2003). To summarize briefly;

The data matrix of the p variable in the n measurement is standardized,

The correlation matrix of the standardized data matrix is found,

The eigen values and standardized eigen vectors of the correlation matrix are calculated,

From the eigenvalues, the explanation ratios of the principal components to the total variance are found,

Principal component values are found by multiplying the transpose of each eigenvector with the standardized data matrix (Ersungur, 2007).

Analysis R software version 4.2.2 (R Core Team, 2022) with *FactoMineR* and *factoextra* packages was used to develop breeding policies by using cluster analysis to group cattle breeds in Samsun according to districts.

RESULTS AND DISCUSSION

According to the results of the clustering analysis applied to examine the distribution of cattle breeds by districts, it was understood that the districts were divided into 5 classes. Looking at Figure 1, it is observed that Bafra, Çarşamba, Tekkeköy and Vezirköprü districts are clustered far from other districts. Therefore, it has been determined that a separate breeding policy can be followed in these

districts compared to other districts. Looking at the hierarchical clustering map, it is understood that the explanatory power is 51.7%.

Hierarchical clustering on the factor map

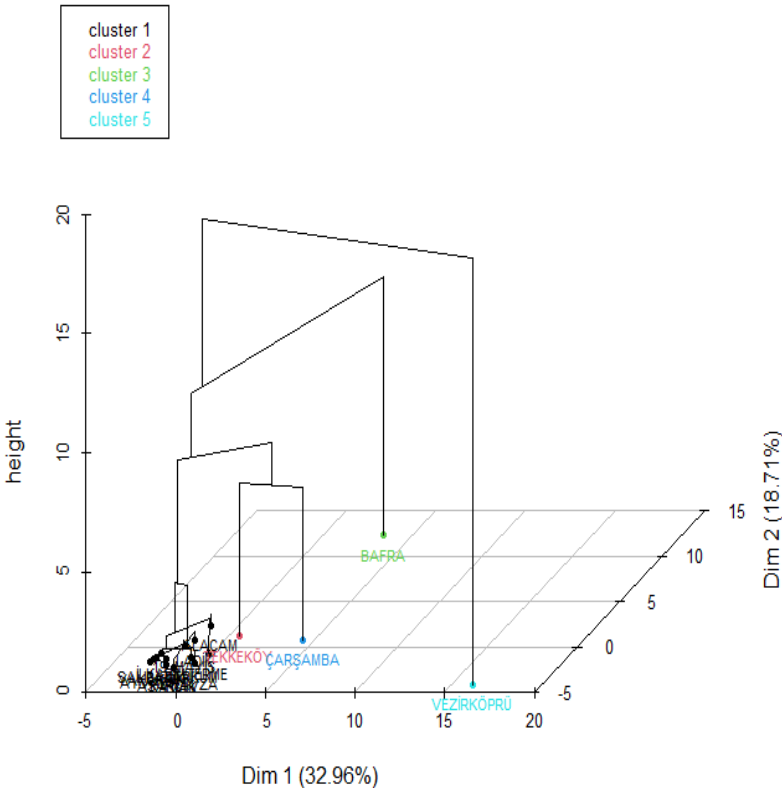


Figure 1. 3D clustering map of districts according to animal breeds.

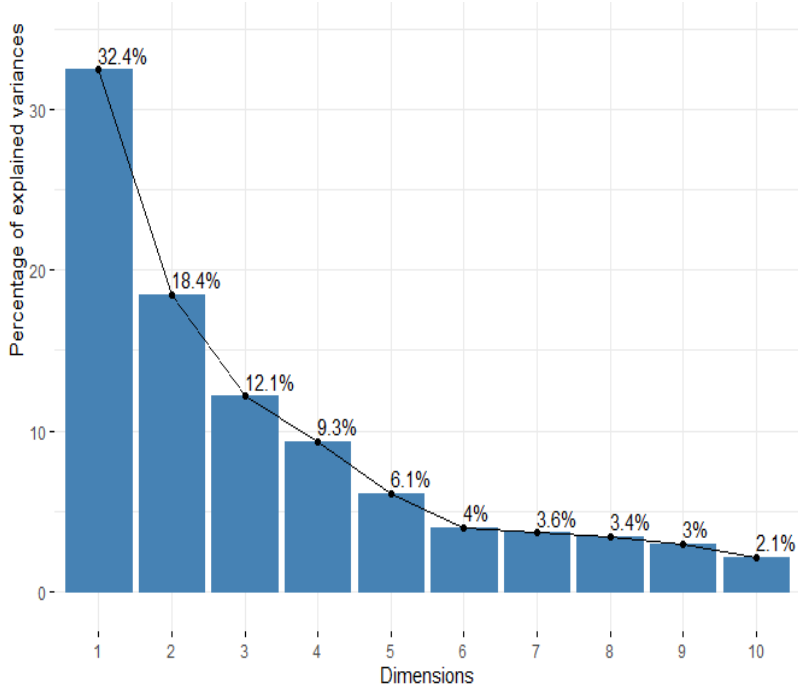


Figure 2. Variance explanatory power of dimensions.

When Figure 2 was evaluated, it was determined that the explanatory power of the dimensions for cluster analysis was 78%.

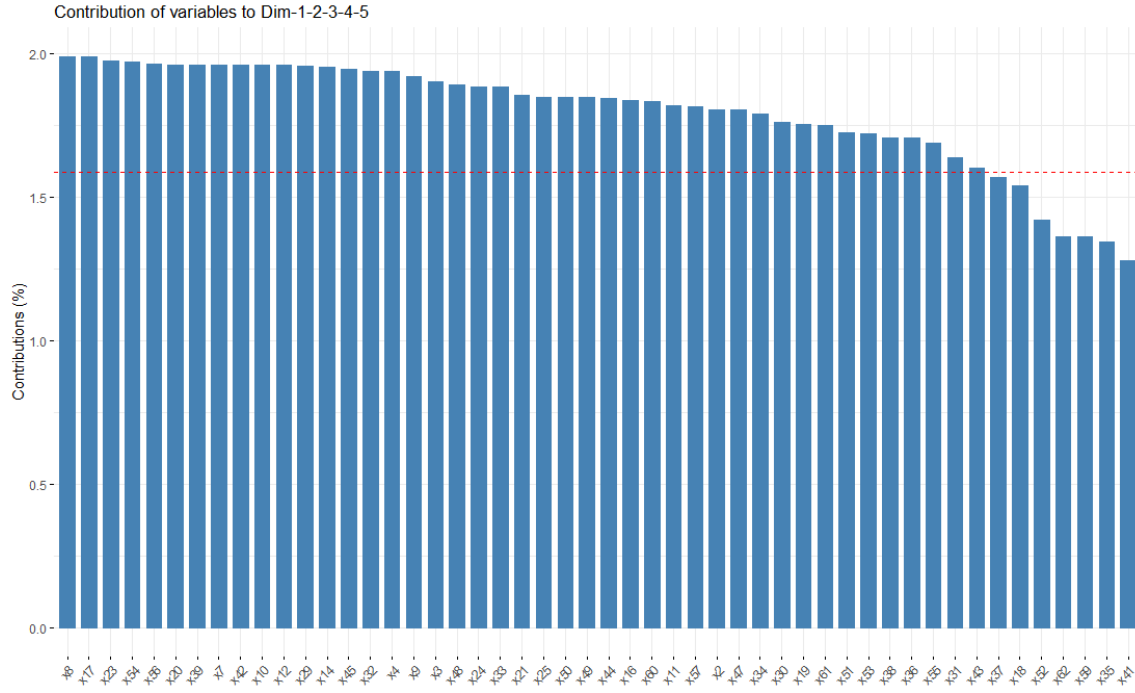


Figure 3. Variables that are significant for data after PCA.

After PCA is done, the variables of the remaining cattle breeds after the insignificant variables in the data set are eliminated are shown in the Figure 3. Animal breeds excluded from the data set in this study; Aberdeen Angus, Aubrac, Avrupa Kırmızısı, Boz Irk Crossbreed, Brangus Melezi, Charalails, Clavel Akman Crossbreed, Estan Kırmızısı, Güney Dogu Anadolu Kırmızısı, Güney Dogu Anadolu Kırmızısı Crossbreed, İsveç Kırmızısı, Jersey, Limosin, Limosin Crossbreed, Normande Crossbreed, Red Angus Crossbreed, Yerli Güney Sarısı, Yerli Güney Sarısı Crossbreed, Zavot, Zavot Crossbreed, Zebu are breeds that were not significant for analysis.

After the variable elimination method was applied, it was observed that all 5 clusters got closer to each other in the new clustering map, when the variables that were meaningless for the data set were removed. It has also been observed that a separate breeding policy can be applied for Bafra, Çarsamba, Tekkeköy and Vezirköprü districts according to the animal breeds raised between the districts. The explanatory power of cluster analysis after PCA increased from 51.7% to 64.1%. Therefore, the applicability of PCA is recommended for data sets with too many variables (Figure 4).

Hierarchical clustering on the factor map

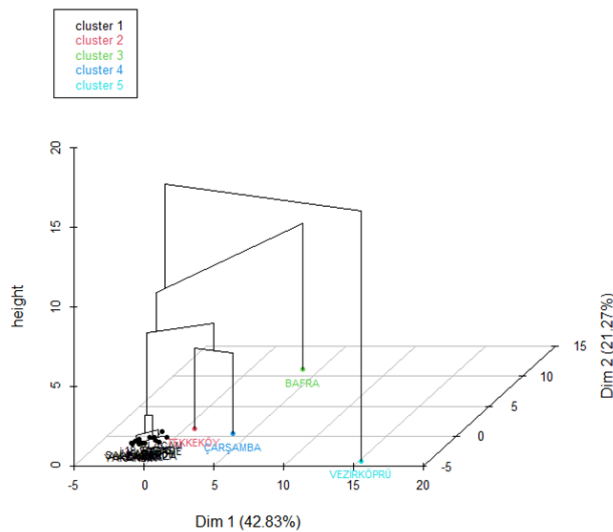


Figure 4. 3D clustering map formed after applying PCA.

CONCLUSION

It has been determined that the same breeding policies cannot be applied in all districts for cattle breeds in Samsun. Accordingly, as a result of the analyzes made, the sustainability of dairy cattle is important for Bafra district, where dairy breeds such as Holstein, Jersey, and Montofon are mostly raised. When we look at Çarşamba county, in addition to dairy breeds, the Yerli Kara Crossbreeds which is a meat breed, is mostly seen, and it may be more suitable for breeding both dairy and meat breeds for this county. Considering the ratio of beef cattle to dairy cattle, it can be said that beef cattle breeds such as Charolais Crossbreeds are suitable for breeding policies as well as dairy breeds for Tekkeköy district. In terms of separate breeding policies, it can be suggested that dairy cattle and beef cattle breeds should be preferred in Vezirköprü district, as the district where the Yerli Kara which is one of the beef cattle breeds, is outnumbered. Traceability of the same cultivation policies is appropriate for other districts, which we can define as cluster 5.

Acknowledgements

This study was published in the proceeding book of VII. International Congress on Domestic Animal Breeding, Genetics and Husbandry - 2023 (ICABGEH-23) Krakow, POLAND, September 18 - 20, 2023 and presented as oral presentation.

REFERENCES

- Abonyi J, Feil B. 2007. Cluster analysis for data mining and system identification. Springer Science, Berlin, Germany.
- Alpar R. 2003. Uygulamalı çok değişkenli yöntemlere giriş 1. Nobel Basımevi, Ankara, Türkiye, pp: 410.
- Aral S, Sakarya E. 1996. Avrupa Birliğine giriş sürecinde Türkiye’de hayvancılık politikaları ve alınması gerekli önlemler. *Avho-Ankara Bölgesi Veteriner Hekimler Odası Dergisi*.
- Bakir G, Ören MY. 2020. Süt sığırcılığı işletmelerinde ırk tercihi ve etkileyen faktörler: Siirt ili örneği. *Kahramanmaraş Sütçü İmam Üniv Tarım Doğa Derg*, 23(5): 1393-1405.
- Çakmak Z. 1999. Kümeleme analizinde geçerlilik problemi ve kümeleme sonuçlarının değerlendirilmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 3: 187-205.
- Çilli M. 2007. İnsan hareketlerinin modellenmesi ve benzesiminde temel bileşenler analizi yönteminin kullanılması. PhD Thesis, Hacettepe University, Institute of Health Sciences, Ankara, Türkiye, pp: 240.
- Demir N. 2012. AB ve Türkiye’de hayvancılık politikalarındaki son gelişmelerin uyum süreci açısından karşılaştırılması. *Alinteri J Agri Sci*, 23(2): 58-63.
- Everitt BS, Landau S, Leese M. 2001. Cluster analysis. Arnold. A member of the Hodder Headline Group, London, UK, pp: 429-438.
- Ersungur ŞM, Kızıltan A, Polat Ö. 2007. Türkiye’de bölgelerin sosyo-ekonomik gelişmişlik sıralaması: Temel Bileşenler Analizi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 21(2): 55-66.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022, URL: <https://www.R-project.org/>.
- Romesburg C. 2004. Cluster analysis for researchers. Lulu, North Carolina, USA.
- Tekin B. 2015. Temel sağlık göstergeleri açısından Türkiye’deki illerin gruplandırılması: Bir kümeleme analizi uygulaması. *Çankırı Karatekin Üniv IIBF Derg*, 5(2): 389-416.
- Timm NH. 2002. Applied multivariate analysis. Springer, New York, US, pp: 296.
- Yılmaz H, Köknaroglu H. 2007. Avrupa birliği ortak tarım politikasına uyum sürecinde Türkiye’de izlenen hayvancılık politikalarının değerlendirilmesi. V. Zootekni Bilim Kongresi, September 5-8, 2007, Erzurum, Türkiye, pp: 11.