

REDUCCIÓN DE LA DIMENSIONALIDAD EN DATOS CRUDOS ESPECTRALES DE LECHE CAPRINA MEDIANTE ANÁLISIS MULTIVARIANTE

DIMENSIONALITY REDUCTION IN RAW SPECTRAL GOAT MILK DATA BY
MULTIVARIATE ANALYSIS

MIRANDA ALEJO, Judith Carmen,
*Instituto Valenciano de Investigaciones Agrarias - Centro de Investigación y Tecnología Animal (CITA-
IVIA)*

PERIS, Cristòfol,
Universitat Politècnica de València

GOMÉZ BLASCO, Ernesto,
*Instituto Valenciano de Investigaciones Agrarias - Centro de Investigación y Tecnología Animal (CITA-
IVIA)*

miranda.scientist@gmail.com
Valencia

Recibido en 14 junio 2023
Aceptado en 23 junio 2023



Resumen

Los investigadores de hoy en día necesitan enfrentarse al análisis de grandes cantidades de datos y requieren software que permita ejecutar diferentes técnicas estadísticas. El análisis de componentes principales (PCA) es una técnica frecuentemente utilizada en análisis exploratorios de datos y simplificación de modelos. En el ámbito de la mejora genética animal, se utilizan análisis químicos de la leche caprina, obtenidos a partir de datos espectrales del infrarrojo medio (MIR) mediante la técnica de transformada de Fourier (FT-MIR). En este estudio, se propone una metodología basada en PCA para seleccionar espectros lecheros específicos, a partir de datos obtenidos de la base de datos de control lechero oficial de la Asociación de Criadores de la cabra Murciano-Granadina de la Comunidad Valenciana (AMURVAL), que comprenden 8742 muestras recolectadas durante los años 2018 y 2021. Los análisis espectrales fueron realizados en LICOVAL con espectrómetros Milkoscan FT6000. En este estudio, se utilizaron datos espectrales crudos (DCE) con 1060 longitudes de onda para realizar un análisis de componentes principales (PCA) utilizando software estadístico R. Se obtuvieron heredabilidades entre 0.00009 y 0.50 en las 1060 longitudes de onda, con una media de 0.17. De las 1060 longitudes de onda, 560 presentaban heredabilidades muy bajas y las restantes 500 tenían heredabilidades medias-altas. En la fase siguiente, se realizó el PCA con las 307 longitudes de onda depuradas, obteniendo 8 componentes principales que explican el 99,5% de la variación total. Los primeros dos componentes principales tienen heredabilidades medias y podrían ser candidatos para programas de mejora animal. En general, la depuración de los datos espectrales y el análisis multivariante permitieron condensar la información en pocas variables, explicando una gran parte de la variación del espectro completo.

Agradecimientos: a AMURVAL-LICOVAL por ceder los DCE. Este trabajo fue posible gracias al contrato postdoctoral financiado por la Generalitat Valenciana y el Fondo Social Europeo (APOSTD/2021/086) y la financiación de los proyectos RTA2017-00049 e IVIA-52201K con financiación FEDER.

Palabras clave: análisis de componentes principales, espectros lecheros, mejora genética.

Abstract

Today's researchers need to handle the analysis of large datasets and require software capable of executing various statistical techniques. Principal Component Analysis (PCA) is a frequently used technique in exploratory data analysis and model simplification. In the field of animal genetic improvement, chemical analyses of goat milk are used, obtained from mid-infrared (MIR) spectral data using Fourier transform (FT-MIR) technology. This study proposes a PCA-based methodology to select specific milk spectra from data in the official milk control database of the Association of Breeders of Murciano-Granadina goats in the Valencian Community (AMURVAL), which includes 8,742 samples collected between 2018 and 2021. Spectral analyses were conducted at LICOVAL using Milkoscan FT6000 spectrometers. Raw spectral data (DCE) with 1,060 wavelengths were used to perform PCA using R statistical software. Heritabilities between 0.00009 and 0.50 were obtained across the 1,060 wavelengths, with an average of 0.17. Of the 1,060 wavelengths, 560 showed very low heritabilities, while the remaining 500 had medium to high heritabilities. In the next phase, PCA was conducted on the 307 cleaned wavelengths, obtaining 8 principal components that explain 99.5% of the total variation. The first two principal components showed medium heritabilities and could be candidates for animal improvement programs. In general, the cleaning of spectral data and multivariate analysis allowed the condensation of information into a few variables, explaining a large portion of the variation in the full spectrum.

Acknowledgments: to AMURVAL-LICOVAL for providing the DCE. This work was made possible by the postdoctoral contract funded by the Generalitat Valenciana and the European Social Fund (APOSTD/2021/086), as well as the funding from projects RTA2017-00049 and IVIA-52201K with FEDER funding.

Key words: principal component analysis, milk spectra, genetic improvement.