



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ZOOTECNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

AVALIAÇÃO DE DESEMPENHO E CARÇA DE OVINOS DA RAÇA SANTA
INÊS: UMA ABORDAGEM MULTIVARIADA

TARLAN OLIVEIRA MILANÊS

RECIFE – PE
JUNHO DE 2019



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ZOOTECNIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

AVALIAÇÃO DE DESEMPENHO E CARÇA DE OVINOS DA RAÇA SANTA
INÊS: UMA ABORDAGEM MULTIVARIADA

TARLAN OLIVEIRA MILANÊS

RECIFE – PE
JUNHO DE 2019

TARLAN OLIVEIRA MILANÊS

AVALIAÇÃO DE DESEMPENHO E CARÇA DE OVINOS DA RAÇA SANTA
INÊS: UMA ABORDAGEM MULTIVARIADA

Dissertação apresentada ao
Programa de Pós-Graduação em
Zootecnia da Universidade
Federal Rural de Pernambuco,
como parte dos requisitos para
obtenção do título de Mestre em
Zootecnia.

Área de concentração: Nutrição e Produção Animal

Comitê de Orientação:

Prof. Dr. Francisco Fernando Ramos de Carvalho (DZ/UFRPE)

Prof.^a Dr.^a Maria Norma Ribeiro (DZ/UFRPE)

Prof.^a Dr.^a Luciana Felizardo Pereira Soares (DZ/UFRPE)

RECIFE – PE

JUNHO DE 2019

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas da UFRPE
Biblioteca Central, Recife-PE, Brasil

M637a Milanês, Tarlan Oliveira.

Avaliação de desempenho e carcaça de ovinos da Raça Santa Inês:
uma abordagem multivariada / Tarlan Oliveira Milanês. - Recife,
2019.

58 f.: il.

Orientador(a): Francisco Fernando Ramos de Carvalho.

Coorientador(a): Maria Norma Ribeiro, Luciana Felizardo Pereira
Soares.

Dissertação (Mestrado) – Universidade Federal Rural de
Pernambuco, Programa
de Pós-Graduação em Zootecnia, Recife, BR-PE, 2019.

Inclui referências.

1. Análise multivariada 2. Carcaça 3. Produção animal
I. Carvalho, Francisco Fernando Ramos de, orient. II. Ribeiro, Maria
Norma, coorient. III. Soares, Luciana Felizardo Pereira, coorient.
IV. Título

CDD 664

TARLAN OLIVEIRA MILANÊS

AVALIAÇÃO DE DESEMPENHO E CARCAÇA DE OVINOS DA RAÇA SANTA
INÊS: UMA ABORDAGEM MULTIVARIADA

Dissertação apresentada e examinada pela comissão de examinadores em 28 de junho de
2019

Orientador:

Prof. Dr. Francisco Fernando Ramos de Carvalho
Universidade Federal Rural de Pernambuco

Comissão Examinadora:

Dra. Janaína Kelli Gomes Arandas
Universidade Federal Rural de Pernambuco

Prof. Dr. João Paulo Ismério dos Santos Monnerat
Universidade Federal Rural de Pernambuco

Prof. Dr. Robson Magno Liberal Veras
Universidade Federal Rural de Pernambuco
(Unidade Acadêmica de Garanhuns)

Dra. Elizabete Cristina da Silva
Universidade Federal Rural de Pernambuco

RECIFE – PE

“Eu só quero o que puder levar”

(Autor desconhecido)

Aos meus pais e noiva.

Dedico

AGRADECIMENTOS

Agradeço, em especial, aos professores que me acompanharam desde o início da graduação, orientando e fazendo perceber o quão especial é essa profissão.

Ao professor Francisco (Chico), por ter me apoiado sempre e por ter me abraçado como orientando.

À professora Luciana Felizardo, por ter me escutado e orientado em todos os momentos que solicitei. Muito orgulho de ter você como professora.

À professora Norma, que desde a graduação me proporcionou a oportunidade de vivenciar momentos muito especiais; pela sua gentileza, pelo poder de mãe e por tudo que me ensinou, principalmente nas horas de desespero. Te agradeço demais.

Aos meus pais, que se dedicaram demais para permitir que eu consiga alcançar meus sonhos.

À minha noiva, que se faz presente sempre que preciso, me aturando nos momentos de estresse e tendo paciência para me direcionar sempre ao melhor caminho.

Às minhas sobrinhas, que me fazem o tio mais feliz do mundo.

A todos que, direta e indiretamente, contribuíram para a conquista de mais um título.

SUMÁRIO

RESUMO GERAL	13
ABSTRACT	15
Considerações Iniciais.....	17
CAPÍTULO I – Referencial Teórico	18
1. Avaliação de Carcaça	20
2. Análise Multivariada.....	21
3. Análise de Componentes Principais.....	22
4. Análise de Agrupamento	25
5. Referências Bibliográficas.....	27
CAPÍTULO II – Análise de Componentes Principais	31
RESUMO	32
ABSTRACT	33
1. Introdução.....	34
2. Material e Métodos	35
3. Resultados e Discussão	37
4. Conclusão	42
5. Referências Bibliográficas.....	43
Capítulo III – Análise de Agrupamento	46
RESUMO	47
ABSTRACT	48
1. Introdução.....	49
2. Material e Métodos	50
3. Resultado e Discussão.....	51
4. Conclusão	56
5. Referências Bibliográficas.....	56
Considerações Finais	58

LISTA DE TABELAS

CAPÍTULO II

Tabela 1 - Médias corrigidas e desvios-padrão das 26 variáveis de consumo e rendimento de carcaça de ovinos da raça Santa Inês.....	37
Tabela 2 - Componentes principais, autovalores e porcentagem de variância.....	37
Tabela 3 - Coeficiente de ponderação (autovetores) dos seis primeiros componentes principais, com capacidade de explicar a variação total das 25 variáveis originais.....	39

CAPÍTULO III

Tabela 1- Distância Euclidiana entre grupos.....	51
Tabela 2 - Grupos com seus respectivos membros (variáveis) e distância para o centro do grupo.....	52
Tabela 3 - Análise de Variância	53

LISTA DE FIGURAS

CAPÍTULO II

Figura 1 - Gráfico bidimensional das variáveis de desempenho e biometria da carcaça de ovinos da raça Santa Inês.

CAPÍTULO III

Figura 2 - Gráfico de médias por grupo.

LISTA DE ABREVIATURAS

CMS - Consumo de Matéria Seca

CMO - Consumo de Matéria Orgânica

CPB - Consumo de Proteína Bruta

CFDN - Consumo de Fibra em Detergente Neutro

CEE - Consumo de Extrato Etéreo

CCHT - Consumo de Carboidratos Totais

CCNF - Consumo de Carboidrato Não Fibroso

CNDT - Consumo de Nutriente Digestível total

NDT - Nutriente Digestível Total

AOL - Área de Olho de Lombo

CP - Componente Principal

Compext - Comprimento Externo

Compint - Comprimento Interno

Larg. Tórax - Largura de Tórax

Pergaru - Perímetro de Garupa

Largaru - Largura de Garupa

Comp. Perna - Comprimento de Perna

Per. Perna - Perímetro de Perna

Prof. Tórax - Profundidade de Tórax

Per. Tórax - Perímetro de Tórax

AVALIAÇÃO DE DESEMPENHO E CARCAÇA DE OVINOS DA RAÇA SANTA INÊS: UMA ABORDAGEM MULTIVARIADA

RESUMO GERAL

Objetivou-se realizar a análise de consumo e rendimento de carcaça de ovinos da raça Santa Inês, por meio da aplicação de técnicas de análise multivariada, visando reduzir a dimensão do espaço multivariado e obter as variáveis de maior importância na formação dos grupos. Foram utilizadas informações de 112 animais, com o intuito de selecionar as variáveis que têm maior capacidade de explicar a variação total dos dados. Os animais do estudo tinham em média cinco meses. Foram selecionadas, a princípio, 26 variáveis oriundas de bancos de dados de experimentos científicos realizados na Universidade Federal Rural de Pernambuco, no Departamento de Zootecnia. Inicialmente, oito variáveis de consumo foram selecionadas: Consumo de Matéria Seca (CMS), Consumo de Matéria Orgânica (CMO), Consumo de Proteína Bruta (CPB), Fibra em Detergente Neutro (CFDN), Consumo de Extrato Etéreo (CEE), Consumo de Carboidratos Totais (CCHT), Consumo de Carboidrato Não Fibroso (CCNF), Consumo de Nutriente Digestível total (CNDT) e Nutriente Digestível Total (NDT), e 17 relacionadas ao rendimento de carcaça, sendo: Peso Médio, Comprimento Externo, Comprimento Interno, Largura do Tórax, Perímetro de Garupa, Largura de Garupa, Comprimento de Perna, Perímetro de Perna, Profundidade do Tórax, Perímetro de Tórax, Peso de Carcaça Fria, Índice de Compacidade da carcaça, Peso Carcaça Quente, Rendimento Carcaça Quente, Rendimento Carcaça Fria, Peso de Carcaça Vazia e Área de olho de lombo. No Capítulo II todas as variáveis foram submetidas à análise de componentes principais, com o objetivo de reduzir a dimensionalidade dos dados. Com a aplicação desta técnica de análise multivariada utilizando como variável de agrupamento o CFDN, foi possível reduzir a variação acumulada dos dados referentes as 25 variáveis restantes, gerando quatro combinações lineares que, juntas, explicam aproximadamente 80% da variação total dos dados, representando cerca de 20% de perda da explicação da variação total. Juntas, as duas primeiras combinações lineares representam 65,07%, sendo possível identificar as variáveis com maior autovetor, em valor absoluto, que foram no CP₁: Peso Carcaça Fria, Peso Carcaça Quente e Peso Corpo vazio (0,27), o que caracteriza esse componente como um índice referente à

conformação de carcaça, visto que todas as variáveis de maior influência estão direcionadas à carcaça. Para o CP₂, as variáveis com maior coeficiente foram: Peso Médio (0,34), Largaru (0,32), RCQ e RCF, ambas com (0,30), indicando um direcionamento deste índice para as características de medidas morfométricas e carcaça. Aplicando o critério de seleção para a escolha das características que tem maior influência em cada componente foram selecionadas variáveis seguindo a ordem de importância do primeiro ao quarto componente, sendo elas: PCF, PCQ, PCVZ todas com (0,27); Peso Médio (0,34); Largura de Garupa (0,32); Comprimento Externo (0,33); NDT (0,34), Profundidade de Tórax (0,33). No Capítulo III, com a aplicação da análise de agrupamento foi possível distribuir as características em quatro grupos distintos. Cada grupo teve um direcionamento, visualizado pela seleção das variáveis que fazem parte de cada um. O grupo 1 tem maior presença de variáveis ligadas a terminação de carcaça. As quatro características que mais contribuíram para formação do grupo foram: PCVZ (Peso de Corpo Vazio), PCF (Peso Carcaça Vazia), PCQ (Peso de Carcaça Quente) e ICC (Índice de Compacidade da Carcaça), descritas em ordem decrescente de importância. No grupo 2 praticamente todas as variáveis contribuíram para a formação do grupo, por apresentarem distância semelhante até o centro do grupo. Esse grupo está relacionado às variáveis de consumo, no qual estão presentes: CMS (Consumo de Matéria Seca), CMO (Consumo de Matéria Orgânica), CPB (Consumo de Proteína Bruta), CCHT (Consumo de Carboidratos Totais) e CNDT (Consumo de Nitrogênio Digestível total). O grupo 3 é relacionado às medidas morfométricas; no entanto, todas as variáveis presentes apresentam certa distância do centro do grupo, tendência que pode indicar a possibilidade de o conjunto de características não representá-lo da melhor forma. No grupo 4 as variáveis com maior influência para formação deste foram: RCF (Rendimento de Carcaça Fria) e RCQ (Rendimento de Carcaça Quente), ambas presentes no segundo componente principal, obtido na análise de componentes principais. A aplicação das técnicas de análise multivariada permitiu identificar, com maior precisão, considerando as relações existentes entre as variáveis, características de maior valor absoluto. Essas, por sua vez, possuem maior capacidade de explicar a variação total dos dados, assim como a maior influência na formação de grupos distintos.

Palavras-chave: Análise Componentes principais. Carcaça. Desempenho. Estatística multivariada

ABSTRACT

The objective of this study was to analyze the consumption and carcass yield of Santa Inês sheep by through the application of multivariate analysis techniques to reduce the size of the multivariate space and obtain the most important variables in the formation of groups. Information from 112 animals was used in order to select the variables that are better able to explain the total variation of the data. The study animals averaged 05 months. At first, 26 variables from databases of scientific experiments conducted at the Federal Rural University of Pernambuco, Department of Zootechnics, were selected. Initially, eight consumption variables were selected: Dry matter consumption (CMS), Organic matter consumption (CMO), Crude protein consumption (CPB), Neutral detergent fiber consumption (CFDN), Ethereal extract consumption Total Carbohydrate Consumption (CCHT), Non-Fibrous Carbohydrate Consumption (CCNF), Total Digestible Nutrient Consumption (CNDT), Total Digestible Nutrient (NDT) and 17 related to carcass yield, being: Average Weight, Length External, Inner Length, Chest Width, Croup Perimeter, Croup Width, Leg Length, Leg Perimeter, Chest Depth, Chest Perimeter, Cold Carcass Weight, Carcass Compass Ratio, Warm Carcass Weight, Carcass Yield Hot, Cold Carcass Yield, Empty Carcass Weight, and Loin Eye Area. In chapter II all variables were submitted to principal component analysis, with the objective of reducing the dimensionality of the data. By applying this multivariate analysis technique using the CFDN as a grouping variable, it was possible to reduce the cumulative variation of data for the remaining 25 variables, generating four linear combinations that together account for approximately 80% of the total variation of the data, representing about 20% loss of explanation of the total variation. Together, the first two linear combinations represent 65.07%, and it is possible to identify the variables with the highest eigenvector, in absolute value, which were in CP1. Cold Carcass Weight, Hot Carcass Weight and Weight Empty body (0.27), which characterizes this component as a carcass conformation index, since all the most influential variables are directed to the carcass. For CP2 the variables with the highest coefficient were: Average Weight (0.34), Largaru (0.32), RCQ and RCF, both with (0.30) indicating a direction of this index to the characteristics of morphometric measurements and carcass. Applying the selection criterion, to choose the characteristics that have the greatest influence on each component, variables were

selected following the order of importance of the first to the fourth component, as follows: PCF, PCQ, PCVZ all with (0.27); Average Weight (0.34); Croup Width (0.32); External Length (0.33); TDN (0.34), Chest Depth (0.33). In Chapter III, with the application of the cluster analysis, it was possible to distribute the characteristics in four different groups. Each group had a targeting, visualized by the selection of the variables that are part of each one. Group 1 has greater presence of variables related to carcass termination. The four characteristics that most contributed to the formation of the group were: PCVZ (Empty Body Weight), PCF (Empty Carcass Weight), PCQ (Hot Carcass Weight) and ICC (Carcass Compaction Index), described in descending order of importance. In group 2, almost all the variables contributed to the formation of the group, because they presented similar distance to the center of the group. This group is related to the consumption variables and the following variables are present: CMS (Consumption of Dry Matter), CMO (Organic Matter Consumption), CPB (Gross Protein Consumption), CCHT (Total Carbohydrate Consumption) and CNDT Total Digestible Nitrogen). Group 3 is related to morphometric measures, however, all variables present in the group present some distance from the center of the group, a trend that may indicate the possibility that the set of characteristics does not represent the group better. In group 04 the variables with the greatest influence for formation of this one were: RCF (Cold Carcass Yield) and WHR (Hot Carcass Yield). Both present in the second main component, obtained in the analysis of main components.

Keywords: Principal componentes analysis. Carcass. Desempenho. Multivariate statistics.

Considerações Iniciais

Segundo dados do IBGE (2017), o efetivo do rebanho de ovinos no Brasil se aproxima das 18 milhões de cabeças. Devido à adaptabilidade da espécie as mais diversas condições ambientais e a habilidade de transformar alimento de baixa qualidade em alimento de alto valor nutritivo, a atividade é praticada em todos os estados do país (SENA et al. 2016).

Os programas de melhoramento de ovinos de corte têm buscado atender as exigências do mercado consumidor, investindo em tecnologias e métodos que facilitem a seleção de animais superiores, capazes de direcionar a maior parte dos nutrientes para a produção de músculos (SANTOS & PÉREZ, 2000). Para minimizar as dificuldades encontradas em algumas regiões, os centros de pesquisa buscam alternativas alimentares, na tentativa de melhorar a produtividade e, com isso, permitir que o setor consiga atender a demanda (ALVES et al, 2014).

Aos produtores resta identificar os pontos limitantes presentes no processo produtivo, de modo que seja possível determinar as melhores estratégias de alimentação e seleção (OSÓRIO et al. 2012). A maioria das pesquisas que envolvem características de carcaça avalia a influência de tratamentos sobre determinadas amostras. As conclusões são, na maioria dos estudos científicos, baseadas em características previamente isoladas, no entanto, os resultados obtidos podem estar incompletos, visto que não estão sendo consideradas as influências exercidas pelos fatores que envolvem todo o processo de produção, além das possíveis correlações existentes entre as características estudadas (GUEDES, 2017).

Os métodos estatísticos multivariados vêm sendo bastante aplicados, graças ao avanço tecnológico, principalmente pela criação de softwares estatísticos destinados a análise das correlações entre as características. As técnicas estão sendo aprimoradas, garantindo informações mais precisas, o que possibilita essencialmente maior confiabilidade nos recursos escolhidos para melhorar os ganhos genéticos e produtivos (MOTTA et al. 2016)

Dessa forma, o objetivo com este trabalho foi aplicar algumas técnicas de análise multivariada, partindo de um banco de informações relacionadas ao desempenho e carcaça de ovinos da raça Santa Inês, com a finalidade de reduzir a dimensionalidade do espaço amostral na análise de componentes principais e identificar as variáveis que melhor explicam a variação total dos dados.

CAPÍTULO I

Referencial Teórico

1. Avaliação de Carcaça

A busca dos consumidores por produtos mais saudáveis direciona o sistema de produção para obtenção de carcaças com peso e idade ao abate ideais (SILVA et al., 2008). A composição das carcaças pode ser estimada por meio da mensuração da área de olho de lombo (AOL) e da espessura da gordura subcutânea. Em muitas situações, os pesquisadores analisam um número maior de características e parte dessas pode ser redundante e/ou contribuir pouco, ou nada, para caracterizar o indivíduo avaliado, sendo essas passíveis de descarte, reduzindo o desgaste do pesquisador (BARBOSA et al., 2005; LIBERATO et al., 1999).

Na ovinocultura, a fase de cordeiro é a categoria que fornece a melhor qualidade de carne, pois, devido à alta velocidade de crescimento, apresenta os melhores rendimentos de carcaça e maior eficiência de produção (PIRES et al., 2000), além de apresentar propriedades benéficas para o organismo humano, principalmente quando provenientes de animais jovens, sendo fonte de proteínas, aminoácidos essenciais e baixa concentração de lipídios e de gorduras saturadas (ALVES et al., 2014).

O crescente mercado consumidor tem estimulado produtores a investir na produção desses animais, principalmente pelo fato de terem alto potencial produtivo (PIRES et al., 2000). Os investimentos no setor seguem alguns critérios, principalmente quando o foco principal dos produtores é produzir carcaças de qualidade, obtendo rendimentos satisfatórios (MOTTA et al., 2016). A seleção é necessária visando animais que possam garantir alta taxa de crescimento, boa reprodução e qualidade de carne, podendo, assim, maximizar os ganhos genéticos por gerações (MAYER, 2017).

Segundo Osório et al. (2012), é necessário conhecer o mercado que quer suprir, no sentido de determinar a melhor forma de alimentar o animal e gerir o tipo de alimento nas diferentes fases de desenvolvimento, fornecendo condições e tempo para seu melhor desenvolvimento. A qualidade da carne vem sendo discutida com mais frequência, devido a maior exigência do mercado. Nesse sentido, é importante ressaltar que, embora seja discutida a qualidade e rendimento de um produto final, esse assunto envolve uma série de fatores prévios ligados ao animal, ao ambiente, ao manejo pré-abate e a gestão pós-abate, fatores esses que possuem o potencial de alterar a estrutura muscular, influenciando diretamente os atributos tecnológicos, visuais e sensoriais (OKEUDO; MOSS, 2005; HOPKINS; MORTIMER, 2014).

As carcaças devem apresentar características favoráveis que contribuam para agregar valor de comercialização (CORDÃO et al., 2012). A conformação e o acabamento de gordura são avaliações qualitativas importantes para indicar a possível quantidade e qualidade de carne vinda da carcaça no sistema de tipificação (COSTA et al., 2012). No entanto, as informações sobre correlações entre as diversas variáveis que tem potencial de afetar a qualidade e rendimento da carcaça ainda necessitam de maiores contribuições literárias (OKEUDO; MOSS, 2005).

Diversos fatores apreciados pelo consumidor podem estar correlacionados aos fatores presentes no sistema de produção (CAMPÊLO, 2015; GIOGO et al., 2016). Nesse sentido, considerando que as características de importância econômica estão correlacionadas com magnitudes e sentidos variáveis, a utilização de técnicas de análise multivariada apresenta a versatilidade de aproveitar as relações existentes (TORRES FILHO et al., 2005; TEIXEIRA NETO et al., 2015).

2. Análise Multivariada

A estatística clássica tem como foco principal a análise isolada de variáveis e, a partir do resultado, realizar considerações relacionadas ao estudo em questão (GUEDES, 2018). No entanto, a variável posta em teste na análise de variância é estendida, sendo acrescentado ao conjunto de dados, múltiplas variáveis dependentes, fazendo com que seja possível considerar a existência de relações. Para suprir essa necessidade de avaliar as inter-relações das variáveis, a alternativa encontrada foi desenvolver metodologias onde todas as variáveis devem ser consideradas ao mesmo tempo. Na análise multivariada as variáveis precisam ser aleatórias e inter-relacionadas de tal maneira que seus possíveis efeitos não podem ser interpretados individualmente (HAIR et al. 2009).

Com o desenvolvimento de metodologias estatísticas que usam de forma simultânea um conjunto de dados, foi possível lidar com a complexidade observada em grande parte dos estudos em relação ao número de variáveis (HAIR, 2009). Os estudos evolutivos iniciaram com Francis Galton em 1886, tendo como multiplicador seu aluno Pearson. Diversos outros pesquisadores passaram a estudar e aplicar as técnicas tais como: Fischer em 1928 e Wilks em 1932. A frequência de uso das técnicas multivariadas teve maior intensificação com o avanço da informática, pois ela

possibilitou a elaboração de softwares estatísticos destinados a essa metodologia (HAIR et al. 2009).

As técnicas multivariadas podem ser utilizadas conforme o objetivo do estudo e o tipo de dados. Kendall (1980) classificou as técnicas como de avaliação interdependente e de avaliação dependente. Com o uso das técnicas interdependentes as variáveis não podem ser classificadas como dependentes ou interdependentes. Todas devem ser analisadas simultaneamente, com o objetivo de encontrar um ponto subjacente a todo o conjunto. Sendo assim, as técnicas que mais se adequam são: análise de componentes principais, correlação canônica, análise de agrupamento e análise de fatores. Já nas técnicas dependentes, a variável pode ser única, possuir diversas variáveis dependentes ou até mesmo diversas relações de dependência, onde as técnicas que mais se adequam são: regressão, relação funcional, múltipla contingência e análise discriminante (HAIR, 2009).

A análise multivariada corresponde a um grande número de métodos e técnicas que utilizam, simultaneamente, todas as variáveis na interpretação teórica do conjunto de dados (NETO, 2004). Com o desenvolvimento tecnológico e descobertas científicas, um grande número de informações pode ser processada antes de ser transformada em conhecimento (VICINI et al. 2005). As técnicas multivariadas têm contribuído significativamente em estudos reprodutivos, de conformação e desempenho animal, de conservação de recursos genéticos, melhoramento genético animal, bem como para um número considerável de áreas de conhecimento (TEIXEIRA NETO et al, 2016).

3. Análise de Componentes Principais

A análise de componentes principais é uma técnica de modelagem da estrutura de covariância. Inicialmente, o objetivo de seu iniciador Pearson, em 1901, foi obter linhas e planos que melhor se ajustassem ao conjunto de pontos em um espaço p-dimensional. A redução do número de variáveis não é feita selecionando determinadas características. Essa redução é realizada pela combinação linear das variáveis iniciais, que vão gerar novas variáveis sintéticas (VICINI et al. 2005). Ou seja, transforma linearmente um conjunto original de variáveis, inicialmente correlacionadas entre si, em combinações lineares, que comportam o maior número possível de informações do

conjunto original com menor perda possível da informação (ABDI & WILLIAMS, 2010; HONGYU et al. 2016).

O objetivo do método é obter informações de um conjunto de dados, expressando essas informações em um novo conjunto de dados, onde cada componente principal gerado é uma combinação linear de todas as variáveis originais para, então, possibilitar a interpretação e explicação da estrutura da variância e covariância (caso as variáveis apresentem distribuição normal) de um vetor aleatório, composto por p -variáveis aleatórias (SOUZA, 2000; SAVEGNANO et al., 2011). A matriz de correlação é utilizada quando a padronização dos dados é necessária, evitando problemas como a influência da magnitude das variáveis. Em outras palavras, deve ocorrer quando, para a determinação dos componentes principais, haja a necessidade de eliminar o efeito de escala das variáveis. Essa alternativa é recomendada quando as variáveis são medidas em escalas diferentes entre si (JOHNSON & WICHERN, 1992).

É ideal que as informações do banco de dados estejam organizadas de acordo com a unidade correspondente e que suas dimensões não sejam tão desproporcionais. Caso contrário, a variável com maior valor tende a ter o maior peso na estimação dos componentes, ou seja, tem maior representatividade devido à unidade, e não pela estrutura das observações, de modo que a padronização é uma ação necessária para evitar possíveis vícios na aplicação do método (MANLY, 2008). Um dos principais usos da análise de componentes principais ocorre quando as variáveis devem ser observadas ao mesmo tempo.

A técnica de componentes principais irá transformar um conjunto de variáveis originais (X_1, X_2, \dots, X_8) em um conjunto de k componentes (Y_1, Y_2, \dots, Y_{17}). Cada componente principal consiste numa combinação linear das variáveis padronizadas, não correlacionadas entre si, e arranjadas em ordem decrescente de variância. As variáveis com maior coeficiente no componente serão aquelas com maior grau de correlação com o componente principal. Para a determinação dos componentes principais é preciso calcular a matriz de variância – covariância, a matriz de correlação, os autovalores, autovetores e, por fim, obter as combinações lineares que vão ser as novas variáveis (REGAZZI, 2001).

Essas combinações lineares representam a seleção de um novo sistema de coordenadas. O novo eixo representa, com a máxima variabilidade, a melhor descrição da estrutura de covariância (JOHNSON & WICHERN, 2007). Os componentes principais dependem da matriz de variância-covariância (Σ) ou da matriz de correlação

(ρ) de X_1, X_2, \dots, X_p . Assim, O vetor $X' = [X_1, X_2, \dots, X_p]$ tem a matriz de covariância Σ com autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Consideremos as seguintes combinações lineares:

$$Y_1 = a_1' X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a_2' X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

...

$$Y_p = a_p' X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

ou

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & a_{q2} & \dots & a_{qp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = AX$$

$(q \times 1)$ $(q \times p)$ $(p \times 1)$

Assim, a combinação linear $Y=AX$ tem:

$$\mu_Y = E(Y) = E(AX) = A\mu_X$$

$$\Sigma_Y = Cov(Y) = Cov(AX) = A\Sigma_X A'$$

Onde μ_X e Σ_X são o vetor da média e a matriz de variância-covariância de X , respectivamente.

Após a combinação, podemos obter:

$$Var(Y_i) = a_i' \Sigma a_i$$

$$Cov(Y_i, Y_k) = a_i' \Sigma a_k$$

Sendo:

i = autovetores, com $i = 1, 2, \dots, p$;

k = componentes principais, com $k = 1, 2, \dots, p$.

O primeiro componente principal retém a máxima variância, o segundo retém a segunda maior variância e é não correlacionado com o primeiro. Dessa forma, o descarte de componentes com variâncias baixas não representa a perda de características, o que ocorre é um rearranjo para ver quais variáveis tem maior representatividade na variabilidade dos dados (HAIR, 2009). As variáveis originais têm a mesma importância estatística, no entanto, os componentes principais tem importância estatística decrescente, sendo possível desprezar os componentes que menos influenciam na variação total dos dados. (NETO, 2004).

Determinar quais são os componentes principais importantes é o principal fator para facilitar a interpretação de dados submetidos à análise multivariada. Qualquer escolha errada pode levar a superestimação ou perda de informações (DRAY, 2008). Existe uma série de métodos para estimar a dimensionalidade do conjunto de dados, sendo aplicado com maior frequência o método de Kaiser (1960), método gráfico proposto por Cattell (1966), também denominado de diagrama *Scree*, e o de Jolliffe (1972).

O método de Kaiser consiste em incluir os componentes cujos valores sejam superiores a 1,0. Já o método de Cattell propõe que o gráfico represente a porcentagem de variação explicada pelo componente nas ordenadas e/ou autovalores em ordem decrescente, de modo que, quando a curva no gráfico passa a ser praticamente paralela ao eixo das abcissas, os componentes que restam são descartados (FERRÉ, 1995; PERES-NETO et al., 2005). Por último, o método de Jolliffe segue os princípios do primeiro método, selecionando os componentes através de seu autovalor, de forma menos rigorosa, considerando como componentes principais os que tenham valor superior a 0,7 (HAIR, 2009).

Além do uso na redução da dimensionalidade, identificação e compreensão dos padrões de associação entre variáveis, a técnica de componentes principais pode ser utilizada para colaborar na identificação da variável de maior prevalência, através do estudo dos coeficientes de correlação entre os componentes e as variáveis originais (SOUSA, 2000; KHATTREE & NAIK, 2000).

4. Análise de Agrupamento

A análise de Cluster é utilizada para criar grupos em função da similaridade existente entre os dados, formando grupos conforme consegue identificar características semelhantes. As semelhanças ou diferenças entre os elementos são definidas a partir de uma função de distância, definida conforme objetivo do estudo (ALMAS, 2015). A análise de cluster é dividida em dois métodos: hierárquico e não hierárquico.

Os métodos hierárquicos utilizam um algoritmo sem uma única partição do conjunto de dados, mas sim os vários agrupamentos possíveis, de modo que o grupo pode ser mesclado a outro em determinado momento. Não exige número inicial de grupos e são classificados em Métodos Aglomerativos, onde todos os elementos vão

sendo agrupados em etapas, um a um, até que se forme um único grupo com todos os elementos e o Método Divisivo, onde todos os elementos começam juntos em um único grupo e vão sendo separados, até que cada elemento seja seu próprio grupo. Os métodos não hierárquicos precisam de uma partição inicial e flexibilidade, uma vez que os elementos podem trocar de grupo durante a execução da análise (LINDEN, 2009).

A análise de Cluster é uma ferramenta que visa à triagem de diferentes elementos presentes em grupos, de modo que a associação existente entre dois elementos do mesmo grupo é máxima, enquanto que o grau de associação entre elementos de grupos diferentes é mínimo. Sendo assim, é um conjunto de técnicas cujo objetivo é agrupar elementos conforme similaridades. No entanto, é importante observar que esses grupos precisam apresentar grande similaridade interna e pouca similaridade externa (ROSE E LEIS, 2002).

Essa análise é uma técnica de interdependência, pois não é possível previamente determinar as variáveis dependentes e independentes. O que ocorre é a análise de interdependência entre todo o conjunto de variáveis. Dessa forma, faz-se necessário uma medida da distância entre eles, sendo os mais próximos mais semelhantes e, conseqüentemente, os mais distantes menos semelhantes. Para medir a diferença do espaço, entre as variáveis pesquisadas, a distância Euclidiana é utilizada para evitar que essa seja influenciada pela escala, sendo necessário que as variáveis sejam todas padronizadas para que possuam a variância de igual unidade (MARCHESAN et al. 2010).

A análise de agrupamento é uma ferramenta útil para analisar dados em diversas situações. Seja reduzindo a dimensão de um conjunto de dados, reduzindo uma série grande de objetos à informação do centro do conjunto, extraindo características escondidas dos dados e desenvolvendo as hipóteses a respeito da sua natureza. Todos os métodos de agrupamento, seja dissimilaridade ou de proximidade, assumem que todos os relacionamentos de importância entre os elementos podem ser descritos por uma matriz. Cada entrada p_{ij} na matriz, é um valor numérico capaz de demonstrar a proximidade existente entre os elementos i e j . Todos os coeficientes de dissimilaridade são funções $d : \Gamma \times \Gamma \Rightarrow \mathfrak{R}$.

Onde Γ é a representação do conjunto de dados. De modo que, essas funções permitem que ocorra a transformação da matriz de dados em uma matriz de distâncias, como representado abaixo (LINDEN, 2009).

$$\Gamma = \begin{bmatrix} x_{11} & \dots & x_{1f} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{if} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nf} & \dots & x_{np} \end{bmatrix} \quad d = \begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \dots & \dots & 0 \end{bmatrix}$$

A representação gráfica na análise de Cluster pode ser feita de várias formas. Usualmente são representados conforme métodos específicos, como no caso do dendograma, associado a métodos hierárquicos. As partições o diagrama de Venn e as tabelas são representações mais genéricas.

5. Referências Bibliográficas

ABDI, H.; WILLIAMS, L.J. Principal component analysis. **WIREs Comp Stat**, v.2 p.433-459, 2010.

ALMAS, S. P. Análise das estatísticas relacionadas ao jogo que discriminam as equipes vencedoras das perdedoras no basquetebol profissional brasileiro. **Revista Brasileira de Educação Física e Esporte**. 29 (4), p. 551- 558, São Paulo, 2015.

ALVES, L. G. C. et al. Produção de carne ovina com foco no consumidor. **Enciclopédia Biosfera**, v. 10, p. 2399- 2415, 2014.

BARBOSA, L.T. et al. Seleção de variáveis de desempenho de suínos por meio da análise de componentes principais. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.57, n.6, p. 805-810, 2005b.

CAMPÊLO, M. C. S. et al. Perfil sanitário e características físico- químicas da carne ovina comercializada *in-natura*. **Revista do Instituto Adolfo Lutz**, 74 (3), p. 207- 215, 2015.

CATTELL, R.B. The screen test for the number of factors. **Multivariate Behavioural Research**, 1, 245-276, 1966.

CORDÃO, M. A. et al. Acabamento de carcaça de ovinos e caprinos – Revisão Bibliográfica. **Agropecuária Científica no Semiárido**, 8 (2), p. 16- 23, 2012.

COSTA, R. G. et al. Utilização de diferentes metodologias para determinação da área de olho de lombo em ovinos. **Archivos de Zootecnia**, 61(236), p. 615 – 618, 2012.

DRAY, S. On the number of principal components: A test of dimensionality based on measurements of similarity between matrices. **Computational Statistics & Data Analysis**. v. 52, p. 2228-2237, 2008.

FERRÉ, L. Selection of components in principal component analysis: a comparison of methods. **Computational Statistics & Data Analysis**, v. 19, p.669-682, 1995.

GIONGO, C.; NALÉRIO, E. S.; BORBA, M. F. S. Características de carcaça e da carne de ovinos da raça Texel com diferentes idades de abate. **XXV Congresso Brasileiro de Ciência e Tecnologia de Alimentos**. P. 6, 2016.

GUEDES, D. G. P.; RIBEIRO, M. N.; CARVALHO, F. F. R. Técnicas multivariadas na análise de características de carcaça de ovinos da raça Morada Nova. **Ciência Rural**, v. 48, n. 9, Santa Maria, 2018.

HAIR, J.R. et al. **Multivariate Data Analysis**. 6ª edição. Upper Saddle River, NJ: Pearson Prentice Hall, 2009.

HONGYU, K. et al. Análise de componentes principais: resumo teórico, aplicação e interpretação. **Engineering and Science**, v.1, ed. 5, p. 83- 90, 2016.

HOPKINS, D. L., MORTIMER, S. I. Effect of genotype, gender and age on sheep meat quality and a case study illustrating integration of knowledge. **Meat Science** (98), 544–555. 2014.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Efetivo da pecuária**. 2017.

JOHNSON, R. A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**, 3th ed, New Jersey, Pretice Hall, 1992.

JOHNSON, R. A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**, 6th ed, Pearson Education, Inc., 2007.

JOLLIFFE, I. T. Discarding variables in a principal component analysis. I Artificial Data. **Applied Statistics**, v.21, p. 160-173, 1972.

KAISER, H. F. The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, v.20, p. 141-151, 1960.

KENDALL, M. G. **Multivariate analysis**. Hodder Arnold, 2 nd edition, 218p. 1980.

KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS software**, Cary: SAS Institute Inc., 2000.

- LIBERATO, J.R.; VALE, F.X.R.; CRUZ, C.D. Técnicas estatísticas de análise multivariada e a necessidade de o fitopatologista conhecê-las. **Fitopatologia Brasileira**, v.24, p.5-8, 1999.
- LINDEN, R. Técnicas de Agrupamento. **Revista Sistema de Informação da FSMA**. N. 4, p.18- 36, 2009.
- MANLY, B.F.J. **Métodos estatísticos multivariados – uma introdução**. 3ª Edição. Porto Alegre: Bookman, 229p., 2008.
- MARCHESAN, T. M. et al. **Avaliação de processo de ensino: uma abordagem multivariada**. 13p. UNIFRA, Santa Maria, 2010.
- MAYER, A. R. **Análise multivariada para características de desempenho, carne e carcaça de uma população multirracial**. Tese (Doutorado em Zootecnia) Universidade Federal de Santa Maria. Rio Grande do Sul, 2017.
- MOTTA, J. F. et al. Desempenho produtivo e qualidade da carcaça de cordeiros avaliados em dois sistemas alimentares. **Boletim de Indústria Animal**, Nova Odessa, v. 73, n. 1, p. 15- 23, 2016.
- NETO, J. M. M. Estatística multivariada. Uma visão didática – metodológica. **Crítica na Rede**. p.1- 13, 2004.
- OKEUDO, N. J.; MOSS, B. W. Interrelationships amongst carcass and meat quality characteristics of sheep. **Meat Science**, (69), p. 1- 8, 2005.
- OSÓRIO, J. C. S et al. Critérios para abate do animal e a qualidade da carne. **Revista Agrarian**, v. 5, p. 433- 443, 2012.
- PERES-NETO, P.R. et al. How many principal components: stopping rules for determining the number of non-trivial axes revisited. **Computational Statistics & Data Analysis**, v. 49, p. 974-997, 2005.
- PIRES, C. C. et al. Crescimento de cordeiros abatidos com diferentes pesos. Constituintes corporais. **Ciência Rural**, v. 30, p. 869- 879, 2000.
- REGAZZI, A. J. **Análise multivariada**. Universidade Federal de Viçosa. Centro de Ciências Exatas e Tecnológicas. Departamento de Informática, 116p. 2001.
- ROSES, C. F.; LEIS, R. P. Um estudo das condições sócio- econômicas de municípios gaúchos através da análise de cluster. FECAP. **Revista Administração On line**, v.3, n. 3, jul/ago/set. 2002.
- SANTOS, C. L.; PÉREZ, J. R. O. **Cortes comerciais de cordeiros Santa Inês. In: ENCONTRO MINEIRO DE OVINOCULTURA**, 1., 2000, Lavras, MG. **Anais...** Lavras: UFLA, p. 149-168, 2000.

SAVEGNANO, R. P. et al. Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population. **Poultry Science**, 90, p. 2174- 2188, 2011.

SILVA, N.V. et al. Características de carcaça e carne ovina: uma abordagem das variáveis metodológicas e fatores de influência. **Acta Veterinaria Brasilica**, v.2, n.4, p.103-110, 2008.

SOUZA, A. M. **Monitoração e ajuste de realimentação em processos produtivos multivariados**. Tese (Doutorado Engenharia de Produção) - Universidade Federal Santa Catarina, 2000.

TEIXEIRA NETO, M. R.; SOUZA, L. E. B. Diversidade fenotípica de linhagens de ovinos Santa Inês por meio de análise multivariada. **Revista Brasileira de Saúde e Produção Animal**. Salvador, v.16, n. 4, 2015.

TEIXEIRA NETO, M. R. Descrição do crescimento de ovinos Santa Inês utilizando modelos não-lienares selecionados por análise multivariada. **Revista Brasileira de Saúde e Produção Animal**, v. 17, n. 1, 2016. Disponível em: <www.rbspa.ufba.br/index.php/rbspa/article/view/article/3413>.

TORRES FILHO, R. A et al. Estudo de divergência genética entre linhas de suínos utilizando técnicas de análise multivariada. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 57, n. 3, p. 390 – 395, 2005.

VICINI, L.; SOUZA, A. M. **Análise multivariada da teoria à prática**. Universidade Federal de Santa Maria, 215p, 2005.

CAPÍTULO II

Análise de Componentes Principais

RESUMO

A análise de componentes principais é uma técnica multivariada amplamente utilizada no processamento de dados. Possui o intuito de reduzir a dimensionalidade do espaço amostral, selecionando variáveis que tenham maior capacidade de explicar a variação total dos dados, além de eliminar as sobreposições. O objetivo deste trabalho foi avaliar o comportamento do banco de dados das características de carcaça de ovinos Santa Inês com base na aplicação da técnica de análise de componentes principais. Foram coletadas informações de 112 ovinos da raça Santa Inês, compreendendo 26 características. Essa técnica foi eficiente por reduzir a variação total em 25 componentes principais (CP), dos quais foram selecionados os quatro primeiros que, juntos, têm a capacidade de explicar aproximadamente 80% da variação total dos dados. Os dois primeiros componentes principais explicam 65,07% dessa variação e as características com maiores coeficientes de ponderação, em valor absoluto no primeiro componente, foram PCF, PCQ, PCVZ (0,27), caracterizando o CP₁ como um índice para determinação da conformação da carcaça. No segundo componente existe maior influência das características de medidas morfométricas, não impedindo que algumas características relacionadas à carcaça tenham coeficiente de ponderação alto. Dessa forma, com a aplicação do método estatístico, foi possível obter informações mais precisas acerca das relações entre as variáveis analisadas, além de identificar as que melhor explicam a variação total dos dados.

Palavras-chave: Combinações lineares. Estatística. Técnicas multivariadas.

ABSTRACT

Principal component analysis is a multivariate technique widely used in data processing. It has the intention of reducing the dimensionality of the sample space, selecting variables that have greater capacity to explain the total variation of the data, besides eliminating the overlaps. The objective of this project was to evaluate the characteristics involved in determining the best carcass yield of growing Santa Inês sheep, using data from 03 scientific experiments that evaluated aspects of consumption and morphometric characteristics. Information was collected from 112 Santa Inês sheep, comprising 26 characteristics. This technique was efficient by reducing the total variation in 25 main components (CP), from which the first four were selected, which together have the capacity to explain about 80% of the total data variation. The first two main components explain 65,07% of this variation and the characteristics with higher weighting coefficients in absolute value in the first component were PCF, PCQ, PCVZ (0.27), characterizing CP1 as an index for conformation determination of the carcass. In the second component there is greater influence of the characteristics of morphometric measures, not preventing some characteristics related to the carcass have a high weighting coefficient. Thus, with the application of the statistical method, it was possible to obtain more accurate information about the relationships between the analyzed variables, as well as identifying the ones that best explain the total variation of the data.

Keywords: Combinations linear. Multivariate techniques. Statistic.

1. Introdução

O avanço da produtividade animal é o principal foco dos produtores, visando sempre atender às demandas de mercado (SALES, 2017). A seleção é necessária visando animais que possam garantir alta taxa de crescimento, boa reprodução, qualidade de carne e satisfatório rendimento de carcaça, maximizando os ganhos genéticos por gerações (MAYER, 2017).

Para que a produção animal seja eficiente, todo o empreendimento agropecuário precisa ser viável e os indicadores zootécnicos servem como instrumentos de acompanhamento, demonstrando os ganhos do produtor (WANDER; MARTINS, 2005). Esses indicadores são importantes para que seja possível realizar a avaliação do potencial produtivo dos animais, possibilitando a melhor escolha do manejo produtivo (MEMÓRIA et al., 2010). As variáveis envolvidas na avaliação desses indicadores têm determinado grau de correlação, sendo necessário analisar o conjunto de dados para que os fatores de maior influência negativa e as sobreposições sejam excluídos. Essa ação busca aumentar e promover a excelência na produção, graças a maior confiabilidade nas informações obtidas pela análise dos dados (LOPES, 2009).

Existe uma preocupação constante na produção de carne, principalmente com aspectos ligados à qualidade, saúde, processamento e a nutrição, que são indicadores importantes do valor global da carcaça (SAÑUDO et al., 2012). No entanto, o consumidor tem procurado produtos que lhe proporcionem segurança, satisfação e praticidade ao consumir (BARBOSA et al., 2006). A produção animal depende de diversos fatores. Um deles é a capacidade que o animal possui de consumir os alimentos em quantidades suficientes para atender às demandas de manutenção e às demandas produtivas. As características da carcaça são importantes no sistema de produção de carne e, apesar dos esforços dos produtores de carne ovina, a maior parte da carne ofertada no Brasil ainda é originada de animais com baixa qualidade de carcaça (RODRIGUES et al., 2015).

Para que as características de rendimento e de carcaça sejam avaliadas é necessário o uso da estatística. Porém, os métodos estatísticos que avaliam uma variável por vez não consideram as interações existentes entre elas e com os demais fatores ligados à produção animal. Esses são, de certo modo, ineficientes, pois não permitem que sejam considerados, no resultado, os efeitos provocados pela ação de outras variáveis (KIRKPATRICK; MEYER, 2004). Por essa razão, foram desenvolvidos

métodos que possibilitam o uso simultâneo de todas as variáveis, permitindo a interpretação teórica do conjunto de dados (GUEDES et al., 2017). Ainda não há um limite prático para o número de características necessárias a serem avaliadas em um experimento; no entanto, é relevante examinar as que poderiam ser desconsideradas, seja por contribuir pouco para explicar a variação dos dados ou por ser consideradas menos discriminantes ou redundantes (CRUZ; REGAZZI, 1997) (FREITAS et al., 2004) (BARBOSA et al., 2006).

Sendo assim, o objetivo deste estudo foi identificar as variáveis com maior poder de explicar a variação total dos dados e eliminar as variáveis que não tenham representatividade e que não contribuam para eleger os principais componentes presentes nos fatores estudados.

2. Material e Métodos

O banco de dados foi obtido a partir das informações de três experimentos científicos realizados na Universidade Federal Rural de Pernambuco, no setor de ovinos. Os dados foram compostos por informações de 112 ovinos da raça Santa Inês, machos, não castrados, com idade média de cinco meses. Do total de dados foram selecionadas 26 variáveis, sendo o Consumo de Fibra Detergente Neutro (CFDN) a variável utilizada para agrupar as demais conforme níveis de FDN. Os níveis de FDN foram divididos em: baixo, entre 14 – 26%, intermediário, entre 27 – 50%, e alto, acima de 51%. As demais foram divididas em dois grupos, sendo oito relacionadas ao consumo: Consumo de Matéria Seca (CMS), Consumo de Matéria Orgânica (CMO), Consumo de Proteína Bruta (CPB), Consumo Extrato Etéreo (CEE), Consumo de Carboidratos Totais (CCHT), Consumo de Carboidrato Não Fibroso (CCNF), Consumo de Nutriente Digestível (CNDT), Nutriente Digestível Total (NDT) e 17 variáveis relacionadas ao rendimento de carcaça, sendo elas: Peso Médio, Comprimento Externo, Comprimento Interno, Largura do Tórax, Perímetro de Garupa, Largura de Garupa, Comprimento de Perna, Perímetro de Perna, Profundidade do Tórax, Perímetro de Tórax, Peso de Carcaça Fria, Índice de Compacidade da carcaça, Peso Carcaça Quente, Rendimento Carcaça Quente, Rendimento Carcaça Fria, Peso de Carcaça Vazia e Área de Olho de Lombo (AOL). Os dados foram submetidos à análise estatística realizada por meio do Software Statistica, Versão 13.3. Todas as variáveis do banco de dados

foram padronizadas pela sua média e desvio padrão, segundo Mingoti (2005). Esta ação teve como objetivo eliminar as diferenças das unidades de medida e obter os componentes principais por meio da matriz de correlação. Pela matriz de correlação foram obtidos os componentes principais, buscando, por meio da comparação, diminuir as diferenças, além de permitir a identificação dos autovetores (JOHNSON; WICHER, 2007).

A princípio, para análise, foi utilizada a matriz de correlação das variáveis X originais: Consumo de Matéria Seca (CMS), Consumo de Matéria Orgânica (CMO), Consumo de Proteína Bruta (CPB), Consumo Extrato Etéreo (CEE), Consumo de Carboidratos Totais (CCHT), Consumo de Carboidrato Não Fibroso (CCNF), Consumo de Nutriente Digestível (CNDT), Nutriente Digestível Total (NDT), Peso Médio, Comprimento Externo, Comprimento Interno, Largura do Tórax, Perímetro de Garupa, Largura de Garupa, Comprimento de Perna, Perímetro de Perna, Profundidade do Tórax, Perímetro de Tórax, Peso de Carcaça Fria, Índice de Compacidade da carcaça, Peso Carcaça Quente, Rendimento Carcaça Quente, Rendimento Carcaça Fria, Peso de Carcaça Vazia e Área de Olho de Lombo (AOL). Com a aplicação da técnica de componentes principais como ponto de partida, a matriz de correlação $P_{25 \times 25}$, do conjunto de variáveis X (X_1, X_2, \dots, X_{25}), foi transformado em um conjunto de k componentes principais (Y_1, Y_2, \dots, Y_k), de modo que cada um consiste em uma combinação linear das variáveis padronizadas, não correlacionadas entre si, e arranjadas em ordem decrescente da variância. Dessa forma, a matriz $P_{25 \times 25}$, pode ser definida por:

$$Y_j = a_{j1}Z_1 + a_{j2}Z_2 + \dots + a_{j25}Z_{25}$$

Em que a_{ij} são os elementos autovetores, com $i=1,2,\dots,25$. A variância de Y_j é igual a $\lambda_j, j=1,2,\dots,25$ e a covariância entre Y_j e Y_k é igual a zero, para qualquer $j \neq k$. A correlação entre o componente Y_j e a variável padronizada Z_i é igual a:

$$r_{Y_j, Z_i} = r_{Y_j, X_i} = a_{ji} / \sqrt{\lambda_j}$$

A importância relativa de um componente foi analisada por meio da porcentagem da variância total, obtida pelo autovalor em relação ao total de autovalores de todos os componentes. A escolha dos componentes que explicarão a variância do conjunto de dados foi determinada pelos autovalores iguais ou superiores a 01 (um), de acordo com o critério de Kaiser (1960). Para análise, o primeiro componente principal agrega a variância máxima entre todas as combinações lineares. Já o segundo componente é não correlacionado com o primeiro e possui a segunda maior variância, ocorrendo o mesmo com os demais (HAIR et al., 2009).

Médias e desvio padrão (DP) das 25 variáveis analisadas estão apresentados na Tabela 1.

Tabela 1. Médias corrigidas e desvios – padrão das 25 variáveis de consumo e rendimento de carcaça de ovinos da raça Santa Inês.

Característica	Média	DP	Característica	Média	DP
AOL (cm ²)	10,32	1,97	Per. Garupa (cm)	59,35	4,28
CMS (Kg)	1,038	0,20	Larg. Garupa (cm)	19,96	3,67
CMO (Kg)	0,951	0,18	Comp. Perna (cm)	40,47	2,11
CPB (Kg)	0,174	0,03	Per. Perna (cm)	36,39	3,16
CEE (g/dia)	25,20	9,71	Prof. Tórax (cm)	25,79	2,50
CCHT (Kg)	0,769	0,14	Per. Tórax (cm)	66,43	3,80
CCNF (Kg)	0,394	0,11	PCF (Kg)	13,77	2,35
CNDT (Kg)	0,763	0,30	ICC (Kg/cm)	0,23	0,03
NDT (%)	71,98	4,17	PCQ (Kg)	14,27	2,37
Peso médio (Kg)	21,99	3,80	RCQ (%)	46,56	3,18
Comp. Ext. (cm)	56,43	2,90	RCF (%)	44,89	3,53
Comp. Int. (cm)	60,32	3,36	PCVZ (Kg)	25,55	4,18
Larg. Tórax (cm)	21,45	1,87			

CMS – consumo de matéria seca, CMO – consumo de matéria orgânica, CPB – consumo de proteína bruta, CEE – extrato etéreo, CCHT – carboidratos totais, CCNF – carboidratos não fibrosos, CNDT – nitrogênio digestível total, NDT – nitrogênio digestível total, peso médio, Comprimento externo, comprimento interno, Largura de Tórax, Perímetro de garupa, Largura de garupa, Comprimento de perna, Perímetro de perna, Profundidade de tórax, Perímetro de tórax, PCF – peso carcaça fria, ICC – índice de compacidade da carcaça, PCQ – peso carcaça quente, RCQ – rendimento carcaça quente, RCF – rendimento carcaça fria, PCVZ – peso de corpo vazio.

3. Resultados e Discussão

Os resultados obtidos com a análise de Componentes Principais estão representados na Tabela 2.

Tabela 2. Componentes principais, autovalores e porcentagem de variância.

CP	Autovalor	% Var. Total	Autovalor acumulado	% Acumulada
01	11,93014	47,72057	11,93014	47,7206
02	4,33826	17,35304	16,26840	65,0736
03	2,16805	8,67221	18,43645	73,7458
04	1,56427	6,25708	20,00072	80,0029

De acordo com as informações, dos 25 componentes principais obtidos, apenas os quatro primeiros foram selecionados (Tabela 2). Foram selecionados os componentes que possuem o maior valor absoluto, dentre os obtidos ou a percentagem da variação total que esse componente explica (KHATTREE; NAIK, 2000; Muniz et al., 2014). Os quatro componentes principais selecionados explicam juntos, aproximadamente, 80% da variação total dos dados, representando cerca de 20% de perda de explicação da variação total.

Com base no critério de seleção de Kaiser (1960), devem ser selecionados os componentes principais com autovalores acima de 01 (um). Dessa forma, a percentagem da variação acumulada é de 80,0029% da variação total dos dados (Tabela 2), com quatro combinações lineares. A determinação dos componentes que vão ser utilizados para fornecer uma interpretação significativa, tanto estatística quanto biológica, é um dos maiores desafios na análise multivariada. No entanto, é de grande importância selecionar componentes que contribuam para eleger as variáveis de maior importância na explicação da variação dos dados (JOLLIFFE, 2002; PERES NETO et al., 2003).

Quando os resultados obtidos são comparados aos dados observados em outros trabalhos que avaliaram características de carcaça e qualidade da carne de ovinos é possível perceber que os resultados são semelhantes. Bonvillani et al. (2010) verificaram que os seis primeiros componentes principais explicaram aproximadamente 85% da variação total para características ligadas à qualidade de carcaça, em seu trabalho de caracterização de carcaças de cabritos da raça Cordobêz. Já Picardi et al. (2010), avaliando carcaças de cordeiros criados a pasto e em confinamento, verificaram que apenas com os dois primeiros componentes era possível explicar 87% da variação total no sistema a pasto e em confinamento.

Santos et al. (2008), avaliando a variação total das medidas de qualidade de carcaça, conseguiram reduzir a dimensionalidade dos dados em quatro combinações lineares das 16 variáveis originais, explicando aproximadamente 80% da variação total. Resultado semelhante foi obtido por Guedes et al. (2017), que conseguiram explicar 80,43% da variação total dos dados com os cinco primeiros componentes principais, mostrando uma redução na dimensionalidade dos dados utilizando o critério de seleção de Kaiser (1960), aplicando as técnicas multivariadas em carcaças de ovinos da raça Morada Nova, o que representa menos de 20% de perda na explicação da variação total.

Caneque et al. (2004) conseguiram explicar 74% da variação total com oito componentes principais quando avaliaram características de qualidade de carcaça.

Os coeficientes de ponderação (autovetores) foram obtidos para cada componente principal (Tabela 3), com base no peso das variáveis avaliadas para estabelecer aquelas com maior grau de importância.

Tabela 3. Coeficiente de ponderação (autovetores) dos quatro primeiros componentes principais, com capacidade de explicar a variação total das 25 variáveis originais.

Variáveis	CP1	CP2	CP3	CP4
AOL	-0,173252	-0,094140	-0,290868	0,084100
CMS	-0,247765	0,162402	0,192116	0,111924
CMO	-0,242332	0,177399	0,181878	0,187320
CPB	-0,240472	0,151840	0,233494	0,075883
CEE	-0,165771	-0,228055	0,184569	0,244229
CCHT	-0,227346	0,215376	0,117524	0,256793
CCNF	-0,210807	-0,145849	0,154656	0,142371
CNDT	-0,243645	0,164954	0,219649	0,184252
NDT	-0,112459	0,056540	0,184956	0,340183
Peso Médio	-0,135652	0,346395	-0,142719	-0,257364
CompExt	-0,175130	0,175530	-0,333317	-0,010263
CompInt	-0,150006	0,251127	-0,346890	0,028987
LargTórax	-0,179822	0,109719	-0,242599	-0,069319
PerGaru	-0,206120	-0,246670	-0,185172	0,039772
LarGaru	-0,080101	-0,327733	-0,277778	0,341746
ComPerna	-0,126087	0,255498	0,010760	-0,207928
PerPerna	-0,169329	-0,290860	-0,209729	0,144162
ProfTórax	-0,119749	-0,026028	0,234648	-0,330416
PerTórax	-0,236286	-0,011271	0,020468	-0,237001
PCF	-0,275642	-0,053858	-0,099695	-0,126466
ICC	-0,254572	-0,154926	0,018481	-0,152035
PCQ	-0,272441	-0,043130	-0,143373	-0,109168
RCQ	-0,138835	-0,302817	0,162293	-0,277149
RCF	-0,139850	-0,302977	0,229648	-0,292425
PCVZ	-0,273921	-0,055355	-0,070978	-0,104645

AOL – Área de olho de lombo, CMS – consumo de matéria seca, CMO – consumo de matéria orgânica, CPB – consumo de proteína bruta, CEE – extrato etéreo, CCHT – carboidratos totais, CCNF – carboidratos não fibrosos, CNDT – nutriente digestível total, NDT – nutriente digestível total, peso médio, Comprimento

externo, comprimento interno, Largura de Tórax, Perímetro de garupa, Largura de garupa, Comprimento de perna, Perímetro de perna, Profundidade de tórax, Perímetro de tórax, PCF – peso carcaça fria, ICC – índice de compactidade da carcaça, PCQ – peso carcaça quente, RCQ – rendimento carcaça quente, RCF – rendimento carcaça fria, PCVZ – peso de corpo vazio.

A variação total nos dois primeiros componentes principais é de 65,07%. Nessas duas combinações lineares estão representadas as características com maior coeficiente de ponderação, em valor absoluto. O primeiro componente (Tabela 3) tem como maiores coeficientes de ponderação as características PCF, PCQ e PCVZ, todas com (0,27), indicando que o índice está relacionado à conformação da carcaça. Já o segundo componente tem como maiores coeficientes de ponderação as características, Peso Médio (0,34), Largaru (0,32) e RCQ (0,30), indicando um direcionamento deste índice para as características de consumo e carcaça.

No segundo componente principal existe uma maior concentração de coeficientes representativos nas características de medidas biométricas. Dessa forma, os dois componentes principais podem ser expressos como os modelos lineares a seguir:

$$CP_1 = AOL (0,17) + CMS (0,24) + CMO (0,24) + CPB (0,24) + CEE (0,16) + CCHT (0,22) + CCNF (0,21) + CNDT (0,24) + NDT (0,11) + \text{Peso Médio} (0,13) + \text{Comp. Ext} (0,17) + \text{Comp. Int.} (0,15) + \text{Larg. Tórax} (0,17) + \text{Per. Garupa} (0,20) + \text{Larg. Garupa} (0,08) + \text{Comp. Perna} (0,12) + \text{Per. Perna} (0,16) + \text{Prof. Tórax} (0,11) + \text{Per. Tórax} (0,23) + PCF (0,27) + ICC (0,25) + PCQ (0,27) + RCQ (0,13) + RCF (0,13) + PCVZ (0,27)$$
$$CP_2 = AOL (0,09) + CMS (0,16) + CMO (0,17) + CPB (0,15) + CEE (0,22) + CCHT (0,21) + CCNF (0,14) + CNDT (0,16) + NDT (0,05) + \text{Peso Médio} (0,34) + \text{Comp. Ext} (0,17) + \text{Comp. Int.} (0,25) + \text{Larg. Tórax} (0,10) + \text{Per. Garupa} (0,24) + \text{Larg. Garupa} (0,32) + \text{Comp. Perna} (0,25) + \text{Per. Perna} (0,29) + \text{Prof. Tórax} (0,02) + \text{Per. Tórax} (0,01) + PCF (0,05) + ICC (0,15) + PCQ (0,04) + RCQ (0,30) + RCF (0,30) + PCVZ (0,05)$$

O bom desempenho produtivo de uma propriedade e o desempenho satisfatório de animais de produção ocorre, na maior parte das vezes, baseado nas informações coletadas e interpretadas. As medidas biométricas e morfométricas, em conjunto com os índices zootécnicos, contribuem para que a caracterização fenotípica de grupos raciais possa auxiliar outros processos quando se busca estabelecer relações entre os aspectos de funcionalidade e conformação, definir o porte e aptidão, além de fazer a seleção

adequada dos grupos que mais contribuirão para que haja um avanço genético (ALMEIDA et al., 2015).

O grupo de variáveis originais dispostas mais afastadas do ponto zero dos dois eixos X e Y, Figura 1, tem maior contribuição para explicar a variação total dos dados.

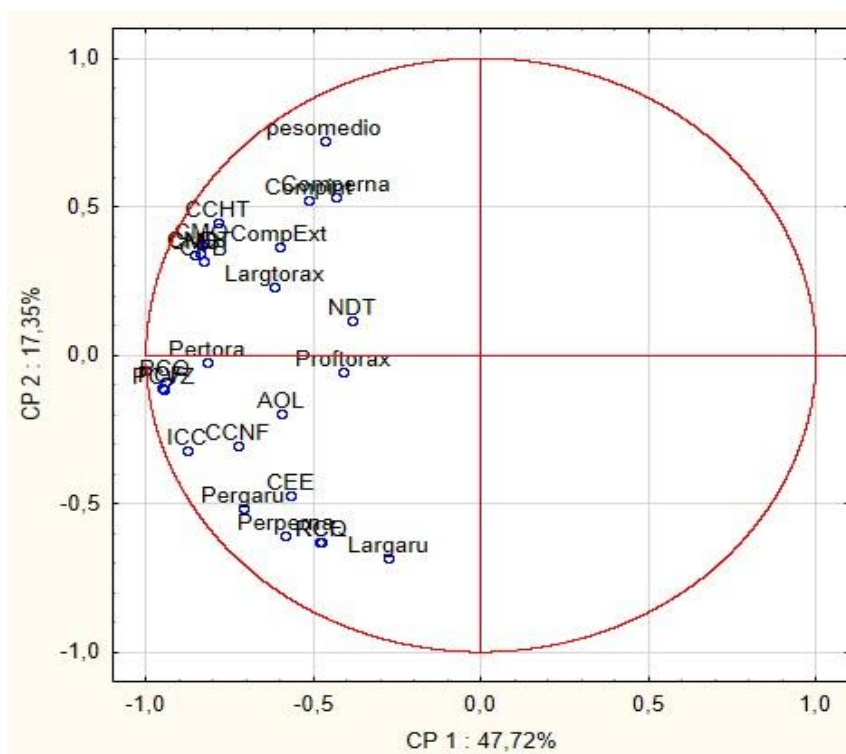


Figura 1. Gráfico bidimensional das variáveis de desempenho e biometria da carcaça de ovinos da raça Santa Inês.

O número de componentes principais selecionados vai direcionar a escolha das variáveis que mais contribuíram para compor o valor absoluto entre as demais (MUNIZ et al. 2014). Esse é um dos principais métodos utilizados, onde são escolhidas as variáveis que mais tiveram representatividade no primeiro componente. Em seguida, o mesmo procedimento é feito para os demais componentes utilizados para explicar a variação total, segundo Khattree & Naik (2000). Sendo assim, utilizando esse critério de seleção, as variáveis que mais contribuíram para a variação total, devendo ser usadas em análises experimentais futuras, são Peso de Carcaça Fria, Peso de Carcaça Quente, Peso de Corpo Vazio (0,27), Peso Médio (0,34), Largaru (0,32), RCF e RCQ, ambas com (0,30). Estas características permitem que algumas medidas corporais possam ser utilizadas como indicadores de crescimento e desenvolvimento do animal (ARAÚJO FILHO et al., 2007; PINHEIRO; JORGE, 2010).

As carcaças consideradas de boa qualidade, segundo Silva Sobrinho (2001), são as que apresentam valores próximos a 14,4 kg, ainda quente, e aproximadamente 13,8 kg já resfriadas. Bezerra et al. (2010) afirmam que o consumidor tem preferência por carcaças de tamanho moderado, entre 12-14 kg, o que determina o abate dos animais entre 28-30 kg. Já Grandis et al. (2016) citam que a faixa de peso para obter deposições de gordura adequadas em ovinos da raça Santa Inês encontra-se entre 15-35 kg pois, acima desse peso, ocorre deposição acentuada de gordura. Ou seja, com base nas informações desses autores, duas variáveis que possuem bastante influência na qualidade de carcaça são Peso de Carcaça Quente e Peso de Carcaça Fria, ambas presentes nas variáveis que mais influenciaram para a variação total.

A conformação do posterior está intimamente ligada à inter-relação existente entre variáveis que estão direcionadas a essa região, tendo melhor conformação quando a correlação entre variáveis de largura e comprimento tem mais proximidade. A deposição de tecido muscular não depende apenas de uma boa nutrição. Para que ocorra uma distribuição proporcional, principalmente em animais confinados, é necessário que haja tempo suficiente para favorecer o desenvolvimento da estrutura torácica do animal a ponto de poder acomodar os órgãos do sistema respiratório e circulatório, em conjunto com o sistema digestivo, sendo possível ter uma circulação de nutrientes capaz de reduzir os déficits (MEDEIROS et al., 2009). Isso melhora não só o desempenho produtivo, tornando adequada a partição desses nutrientes, fato esse que contribui para uma melhor deposição muscular e para uma melhor conformação da carcaça.

A diminuição do conjunto de variáveis permite que as medidas corporais possam ser simplificadas em poucas variáveis, facilitando a seleção e reduzindo os custos (SILVA, 2013). O conhecimento das relações das características fornece importantes informações para a escolha de programas de seleção, visando a uma alta eficiência no melhoramento das características (SALES, 2017).

4. Conclusão

O avanço tecnológico contribuiu muito para que fosse possível interpretar e relacionar banco de dados. Os programas estatísticos têm papel importante, facilitando o melhor entendimento, além de possibilitar um direcionamento mais preciso na tomada de decisões. A aplicação da técnica de componentes principais contribuiu para a identificação de variáveis importantes, tanto para desenvolvimento do animal quanto para a obtenção de carcaças com maior rendimento, uma vez que, pela interpretação dos

resultados foi possível identificar as características que possuem maior capacidade de explicar a variação total dos dados.

5. Referências Bibliográficas

ALMEIDA, J. C. S. et al. Desempenho, medidas corporais, rendimentos de carcaça e cortes, e qualidade de carne em cordeiros alimentados com resíduos da agroindústria processadora de frutas. **Semina: Ciências Agrárias**, Londrina, v.36, n.1, p.541-556, jan/fev, 2015.

ARAÚJO FILHO, J. T. et al. Efeito de dieta e genótipo sobre medidas morfométricas e não constituintes da carcaça de cordeiros deslanados terminados em confinamento. **Revista Brasileira de Saúde e Produção Animal**, v.8, n.4, p. 394-404, out/dez, 2007.

BARBOSA et al. Avaliação de características de qualidade da carne de suínos por meio de componentes principais. **Revista Brasileira de Zootecnia**, v.35, n.4, p. 1639-1645, 2006.

BEZERRA, L. R. et al. Desempenho de cordeiros Santa Inês submetidos a aleitamento artificial enriquecido com *Spirulina plantensis*. **Ciência Animal Brasileira**, 11 (2), p. 258- 263, 2010.

BONVILLANI, A. et al. Carcass characteristics of Criollo Cordobés kid goats under an extensive management system: Effects of gender and liveweight at slaughter. **Meat Science**, v.86, p.651-659, 2010.

CAÑEQUE, V. et al. Carcass and meat quality of light lambs using principal component analysis. **Meat Science**, v.67, p.595-605, 2004.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. 2 ed. Viçosa, MG: Universidade Federal de Viçosa, 1997. 390p.

FREITAS, R. T. F. et al. Avaliação de carcaças de suínos da raça Large White utilizando medidas convencionais. **Revista Brasileira de Zootecnia**, v.33, p. 2037-2043, 2004.

GUEDES, D. G. P. et al. **Técnicas estatísticas multivariadas aplicadas à caracterização de carcaça de ovinos da raça Morada Nova**. Recife, PE: UFRPE, 2017, 99P. Tese (Doutorado Integrado em Zootecnia) - Universidade Federal Rural de Pernambuco, Universidade Federal da Paraíba e Universidade Federal do Ceará, 2017.

GUEDES, D. G. P.; RIBEIRO, M. N.; CARVALHO, F. F. R. Técnicas multivariadas na análise de características de carcaça de ovinos da raça Morada Nova. **Ciência Rural**, v. 48, n. 9, Santa Maria, 2018.

GRANDIS, F. A. et al. Características de carcaça e qualidade da carne de cordeiros alimentados com diferentes teores de tota de soja em substituição ao farelo de soja. **Ciência Animal Rural**, 17, p. 327 – 341, 2016.

JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**, 6th ed, Pearson Education, Inc., 2007.

JOLLIFFE, I. T. Discarding variables in a principal component analysis. I Artificial Data. **Applied Statistics**, v.21, p. 160-173, 1972.

_____. Discarding variables in a principal component analysis. II Real Data. **Applied Statistics**, v. 22, p. 21-31, 1973.

_____. **Principal Component Analysis**. 2ed. 2002.

HAIR, J.R. et al. **Multivariate Data Analysis**. 6ª edição. Upper Saddle River, NJ: Pearson Prentice Hall, 2009.

KAISER, H. F. The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, v.20, p.141-151, 1960.

KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS software**, Cary: SAS Institute Inc., 2000.

KIRKPATRICK, M.; MEYER, K. Direct estimation of genetic principal componentes: simplified analysis of complex phenotypes. **Genetics**. V. 168, p. 2295-2306, 2004.

LOPES, M. A. et al. Influência de diferentes índices zootécnicos na composição e evolução de rebanhos bovinos leiteiros. **Ciência Animal Brasileira**, Samambaia, v. 10, n 2, p. 446-453, 2009.

MAYER, A. R. **Análise multivariada para características de desempenho, carne e carcaça de uma população multirracial**. Tese (Doutorado em Zootecnia) Universidade Federal de Santa Maria. Rio Grande do Sul, 2017.

MEDEIROS, G. R. et al. Efeito dos níveis de concentrado sobre as características de carcaça de ovinos Morada Nova em confinamento. **Revista Brasileira de Zootecnia**, v.38, n.4, p.718-727, 2009.

MEMÓRIA, H. Q. et al. Indicadores zootécnicos de ovinos criados em diferentes sistemas de produção na região norte do Ceará. In: **Ciência e Tecnologia na Pecuária de caprinos e ovinos**. Fortaleza, p. 295-310, 2010.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

MUNIZ, C. A. S. D. et al. Análise de componentes principais para características de crescimento em bovinos de corte. **Ciências Agrárias**, Londrina, v. 35, n. 3, p. 1569-1579, 2014.

PERES-NETO, P.R.; JACKSON, D.A.; SOMERS, K.M. Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. **Ecology** **84**, 2347–2363. 2003.

PICARDI, L. A. et al. Verifying carcass trait changes in a backcross programme with Texel Breed. **Livestock Science**, v.127, p.267-271, 2010.

PINHEIRO, R. S. B.; JORGE, A.M. Medidas biométricas obtidas in vivo e na carcaça de ovelhas de descarte em diferentes estágios fisiológicos. **Revista Brasileira de Zootecnia**, v.39, n.2, p.440-445, 2010.

RODRIGUES, L. S. et al. Características da carcaça e da carne de vacas de descarte abatidas com distintos pesos e grau de acabamento – abordagem meta-analítica. **Cienc. anim. bras.**, Goiânia, v.16, n.4, p. 508-516 out./dez. 2015.

SALES, L. H. B. **Análise de fatores e componentes principais genéticos para características de crescimento, carcaça e qualidade da carne em bovinos da raça Nelore**. Dissertação (Mestrado em Zootecnia), Faculdade de Ciências Agrárias e Veterinárias. Universidade Estadual Paulista. São Paulo, p.51, 2017.

SAÑUDO, C. et al. Características de carcaça e qualidade da carne de lactentes e cordeiros. **Revista Espanhola de Pesquisa Agrícola**, v. 10, n. 3, p. 690-700, 2012.

SANTOS, V.A.C. et al. The use of multivariate analysis to characterize carcass and meat quality of goat kids protected by the PGI “Cabrito de Barroso”. **Livestock Science**, v.116, p.70-81, 2008.

SILVA SOBRINHO, A. G. Aspectos quantitativos e qualitativos da produção de carne ovina. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 41, Piracicaba. **Anais...** Piracicaba: Fundação de Estudos Agrários Luiz de Queiroz. p.425-446. 2001.

SILVA, M.S. **Análise de componentes principais de características mensuradas em ovinos Morada Nova participantes de testes de desempenho**. Dissertação (Mestrado em Produção Animal), Universidade Federal Rural do Semi-Árido, Mossoró, RN, p. 51. 2013.

Capítulo III

Análise de Agrupamento

RESUMO

O estudo na busca por padrões de semelhança pode ser realizado pela formação de grupos por similaridade. A análise de agrupamento estuda todo um conjunto de relações interdependentes, não fazendo distinção entre variáveis. Dessa forma, o objetivo deste trabalho foi identificar as características mais importantes para formação de cada grupo e, conseqüentemente, para a caracterização de carcaças de qualidade em ovinos Santa Inês. Para a análise foram utilizados dados de três experimentos científicos que avaliaram aspectos de consumo e características morfométricas. Foram coletadas informações de 112 ovinos da raça Santa Inês, compreendendo 26 características. Com a análise, foi possível identificar as características de maior contribuição para a formação dos grupos e para agregar novas características a estes. No grupo 1, as características mais importantes foram: PCVZ, PCF, PCQ e ICC, tendo maior predominância das variáveis de carcaça. Já no grupo 2 se concentraram as características relacionadas ao desempenho, sendo as mais importantes: CMS, CMO, CPB, CCHT e CNDT. No grupo 3, as características tinham relação com as medidas morfométricas, muito embora todas tenham apresentado certa distância do centro do grupo. Por fim, no grupo 4, concentraram-se as características relacionadas ao rendimento da carcaça, sendo as de maior relevância: RCF e RCQ. A identificação das variáveis com maior poder de atração para os grupos permite que seja possível direcionar melhor os estudos desta natureza, isto é possível mediante ao maior foco do pesquisador em características importantes para explicar a formação dos grupos.

Palavras-chave: Análise de dados. Carcaça. Santa Inês.

ABSTRACT

The study in the search for patterns of resemblance can be done through the formation of groups by similarity. A cluster analysis studies a whole set of interdependent relationships, not distinguishing between variables. In this way, the group was characterized as an important characteristic for the formation of each group and, consequently, for characterization of quality. For analysis, we used data from 03 scientific experiments that evaluated aspects of consumption and morphometric characteristics. Data were collected from 112 Santa Inês sheep, comprising 26 characteristics. With the analysis, it was possible to identify the characteristics of greater contribution to the formation of the groups and to add new characteristics to them. In group 1, the most important characteristics were: PCVZ, PCF, PCQ and ICC. In group 2, the performance-related characteristics were concentrated, being the most important: CMS, CMO, CPB, CCHT and CNDT. In group 3, the characteristics were related to the morphometric measures, although all had some distance from the center of the group. Finally, in group 4, the characteristics related to the carcass yield were concentrated, being the most relevant: RCF and RCQ. The identification of the most attractive variables for the groups allows to better direct the studies of this nature, this is possible through the researcher's greater focus on important characteristics to explain the formation of the groups.

Keywords: Carcass. Data analysis. Santa Inês.

1. Introdução

A análise de agrupamento estuda todo um conjunto de relações interdependentes, não fazendo distinção entre variáveis, como na regressão (MALHOTRA, 2006). A finalidade da técnica é reunir elementos observados, com homogeneidade em um grupo e com heterogeneidade entre outros (SEIDEL et al., 2008). Os agrupamentos ocorrem devido ao fato de os objetos serem relativamente semelhantes aos presentes no mesmo grupo e diferentes de grupos aleatórios. A análise de agrupamento é o inverso da análise de fatores, pois reduz o número de objetos, mas não o número de variáveis, concentrando em um número menor de grupos (REGAZZI, 2001).

Essa técnica computacional tem o propósito de separar elementos em grupos, considerando suas características (SEIDEL et al., 2008). A maior parte das análises de agrupamento realizadas tem como finalidade tratar a heterogeneidade dos dados (LATTIN, 2011). Basicamente seria como reunir vários elementos com características semelhantes pré-determinadas, em um mesmo grupo. Esses grupos devem seguir uma métrica de qualidade, sendo altamente homogêneos internamente e heterogêneos entre os grupos (HAIR et al., 2009). Cada elemento representa um acréscimo de dados e pode ser constituído por um campo numérico ou categórico (GORDON, 1981).

A análise de Cluster é utilizada para criar grupos em função da similaridade existente entre os dados, formando grupos conforme consegue identificar características semelhantes. As semelhanças ou diferenças entre os elementos são definidas a partir de uma função de distância, definida conforme objetivo do estudo (ALMAS, 2015).

Com essa análise é possível reduzir a dimensão de um conjunto de dados, reduzir um número considerável de elementos em informações e extrair características escondidas dos dados, podendo desenvolver hipóteses acerca do que está sendo analisado (NETO, 2016). É uma técnica de mineração de dados multivariados e das informações das variáveis de cada caso, consegue agrupar automaticamente os casos da base de dados em grupos. Ao agrupar dados similares é possível descrever de forma coerente as características de cada grupo identificado, fornecendo mais confiabilidade no entendimento do conjunto de dados original, o desenvolvimento de novos métodos de classificação para novos dados, além de ter conhecimento de correlações que possivelmente não seriam visualizadas (HAIR, 1995; HAIR, 2009).

A clusterização, como também é conhecida à análise de Cluster, popularizou-se devido à grande quantidade de dados coletados nas diferentes áreas de conhecimento. São comuns aplicações direcionadas ao reconhecimento de padrões e análise de dados, sendo imprescindível cumprir ao menos um dos objetivos principais, identificar estruturas a fim de gerar hipóteses, identificar características marcantes e permitir a classificação natural, com o intuito de identificar o grau de semelhança e facilitar a compreensão por meio dos grupos formados (LINDEN, 2009).

A distância euclidiana é a medida de distância mais usual, quando as variáveis são quantitativas, de modo que a distância entre os indivíduos P e Q, sendo p_1, p_2, \dots, p_n e q_1, q_2, \dots, q_n , os registros dos elementos, são obtidas por:

$$d_{pq} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Na busca contínua dos produtores por animais cada vez mais produtivos, todos os meios possíveis para auxiliar as decisões contribuem para a seleção de indivíduos superiores. Sendo assim, o estudo aplicado a animais de uma população por meio da análise de agrupamento permite a identificação de grupos superiores para as características desejadas (CRUZ et al., 2016).

A primeira etapa da análise de agrupamento é determinar qual medida de similaridade será aplicada para determinar o número real de grupos existentes na amostra. A segunda é a escolha da análise de agrupamento a ser aplicada para agrupar os animais e formar os grupos. Na terceira e última será traçado o perfil dos elementos presentes em cada grupo, para então determinar a sua composição (MALHORTA, 2006; HAIR et al., 2009; COSTA, 2019).

2. Material e Métodos

O banco de dados foi obtido a partir das informações de três experimentos científicos realizados na Universidade Federal Rural de Pernambuco, no setor de ovinos. Os dados foram compostos por informações de 112 ovinos da raça Santa Inês, machos, não castrados. Do total de dados, foram selecionadas 26 variáveis; dessas, nove relacionadas ao consumo, sendo elas: Consumo de Matéria Seca (CMS), Consumo de Matéria Orgânica (CMO), Consumo de Proteína Bruta (CPB), Consumo Fibra Detergente Neutro (CFDN), Consumo Extrato Etéreo (CEE), Consumo de Carboidratos

Totais (CCHT), Consumo de Carboidrato Não Fibroso (CCNF), Consumo de Nutriente Digestível (CNDT) e Nutriente Digestível Total (NDT), e 17 variáveis relacionadas ao rendimento de carcaça, sendo elas: Peso Médio, Comprimento Externo, Comprimento Interno, Largura do Tórax, Perímetro de Garupa, Largura de Garupa, Comprimento de Perna, Perímetro de Perna, Profundidade do Tórax, Perímetro de Tórax, Peso de Carcaça Fria, Índice de Compacidade da carcaça, Peso Carcaça Quente, Rendimento Carcaça Quente, Rendimento Carcaça Fria, Peso de Carcaça Vazia e Área de Olho de Lombo (AOL). Os dados foram submetidos à análise estatística realizada por meio do Software Statistica (versão 13.3).

As variáveis foram submetidas à padronização por meio do procedimento Standardize, do Software Statistica (versão 13.3), antes das distâncias serem calculadas, de modo que todas as variáveis têm igual importância na determinação destas distâncias. Assim, torna-se possível obter a medição de similaridade. Após a padronização, para formar os grupos, foi utilizado o método de agrupamento K-médias para distribuir as características conforme proximidade, dentre os grupos pré-selecionados. Pelo método de k-médias é necessário estabelecer um número inicial de grupos; desta forma, foram formados quatro grupos (clusters) com as características disponíveis para distribuir as variáveis foi calculada a distância média de cada característica. Com isso, foi possível identificar o direcionamento das variáveis mais significativas para caracterizar a carcaça de ovinos.

3. Resultado e Discussão

As distâncias entre os grupos formados estão apresentadas na Tabela 1.

Tabela 1. Distância euclidiana entre grupos

Grupo	Distância Euclidiana (entre grupos)			
	1	2	3	4
1	0,000000	0,602420	0,827169	0,473270
2	0,776157	0,000000	0,535008	1,007083
3	0,909488	0,731442	0,000000	1,358496
4	0,687946	1,003535	1,165545	0,000000

A análise de agrupamento tem como princípio que cada observação de uma amostra pode ser entendida como um ponto em um espaço euclidiano multidimensional. A distância entre duas observações indica a similaridade existente entre as observações (MORAES, 2016).

Com a aplicação da análise de Cluster, pela técnica de agrupamento K-Médias, foi possível formar quatro conjuntos, distribuindo as 26 variáveis selecionadas para o estudo, com a maior distinção possível entre eles. Computacionalmente, esse método de análise de variância começará com k-conjuntos aleatórios e moverá as variáveis entre eles. O objetivo é minimizar a variabilidade dentro dos conjuntos e maximizar a variabilidade entre conjuntos (HAIR, 2009). A ação do programa de mover as variáveis entre os grupos é realizada para obter resultados mais significativos (MALHORTA, 2006).

Tabela 2. Grupos com seus respectivos membros (variáveis) e distância para o centro do grupo.

Variáveis distribuídas de acordo com o grupo e distância do centro do grupo							
GRUPO 01		GRUPO 02		GRUPO 03		GRUPO 04	
Variável	Distância	Variável	Distância	Variável	Distância	Variável	Distância
AOL	0,752200	CMS	0,268626	CFDN	0,759511	Prof. tór	0,771460
CEE	0,710324	CMO	0,208391	Pes. Méd	0,543326	RCQ	0,439932
CCNF	0,660637	CPB	0,329149	Comp. ext	0,581675	RCF	0,354515
Per.garup	0,467377	CCHT	0,278311	Comp. Int	0,508312		
Lar.garup	0,843426	CNDT	0,155228	Larg. Tór	0,719966		
Per. Pern	0,586518	NDT	0,860332	Comp.pern	0,745365		
Per. tórax	0,644097						
PCF	0,387133						
ICC	0,390227						
PCQ	0,410598						
PCVZ	0,396770						

AOL – Área de olho de lombo, CMS – consumo de matéria seca, CMO – consumo de matéria orgânica, CPB – consumo de proteína bruta, CFDN – consumo de fibra detergente neutro, CEE – extrato etéreo, CCHT – carboidratos totais, CCNF – carboidratos não fibrosos, CNDT – nutriente digestível total, NDT – nutriente digestível total, peso médio, Comprimento externo, comprimento interno, Largura de Tórax, Perímetro de garupa, Largura de garupa, Comprimento de perna, Perímetro de perna, Profundidade de tórax, Perímetro de tórax, PCF – peso carcaça fria, ICC – índice de compactidade da carcaça, PCQ – peso carcaça quente, RCQ – rendimento carcaça quente, RCF – rendimento carcaça fria, PCVZ – peso de corpo vazio.

Conforme Tabela 2, as variáveis que menores distâncias apresentaram do centro do grupo, valor demonstrado na seção (GRUPO 1), pela coluna distância, são as que maior influência tem na determinação das variáveis agregadas ao grupo 1. Considerando os níveis de significância do teste F, conforme Tabela 3, as variáveis mais importantes

para a atribuição de objetos ao primeiro grupo foram: PCVZ (Peso de Corpo Vazio), PCF (Peso Carcaça Vazia), PCQ (Peso de Carcaça Quente) e ICC (Índice de Compacidade da Carcaça), descritas em ordem decrescente de importância.

Tabela 3. Análise de Variância

Variáveis	Between group	Within group	Teste F
GFDN	2,71375	97,47546	2,6170
AOL	23,92532	75,46547	29,8014
CMS	44,69072	45,31595	92,7031
CMO	41,06084	48,35543	79,8198
CPB	38,62941	45,17685	80,3767
CEE	15,67595	71,89348	20,4962
CCHT	38,09417	55,13358	64,9487
CCNF	33,59381	62,77657	50,3025
CNDT	42,34248	52,65752	75,5864
NDT	5,60136	89,39864	5,8897
Pesomédio	21,29558	86,50067	23,1418
CompExt	31,29162	58,95181	49,8952
CompInt	27,50705	63,83252	40,5070
LargTórax	32,37367	64,56802	47,1305
PerGaru	32,33304	63,53461	47,8370
LarGaru	4,29258	96,78779	4,1689
CompPerna	20,12189	84,23705	22,4540
PerPerna	22,95934	76,68808	28,1423
ProfTórax	7,41664	90,43102	7,7093
PerTórax	36,74906	50,77930	68,0280
PCF	56,28687	32,88640	160,8861
ICC	40,34931	43,27430	87,6464
PCQ	56,36958	34,64620	152,9386
RCQ	8,05737	76,54351	9,8949
RCF	8,05080	72,37325	10,4566
PCVZ	58,48207	32,99821	166,5943

AOL – Área de olho de lombo, CMS – consumo de matéria seca, CMO – consumo de matéria orgânica, CPB – consumo de proteína bruta, CFDN – consumo de fibra detergente neutro, CEE – extrato etéreo, CCHT – carboidratos totais, CCNF – carboidratos não fibrosos, CNDT – nutriente digestível total, NDT – nutriente digestível total, peso médio, Comprimento externo, comprimento interno, Largura de Tórax, Perímetro de garupa, Largura de garupa, Comprimento de perna, Perímetro de perna, Profundidade de tórax, Perímetro de tórax, PCF – peso carcaça fria, ICC – índice de compacidade da carcaça, PCQ – peso carcaça quente, RCQ – rendimento carcaça quente, RCF – rendimento carcaça fria, PCVZ – peso de corpo vazio.

As características destacadas no grupo I, Tabela 2, são variáveis direcionadas à carcaça, caracterizando este grupo com variáveis que explicam melhor a conformação desta. As técnicas de análise multivariada são usadas em conjunto, com intuito principal de reafirmar os resultados obtidos com sua aplicação. O primeiro componente principal e, conseqüentemente, o de maior influência na explicação da variação total, na análise

de componentes principais, teve como principais variáveis PCF, PCQ e PCVZ, demonstrando a importância que essas variáveis tiveram nas duas análises realizadas.

A possibilidade de cruzar informações das duas análises, utilizando um mesmo banco de dados, enaltece a importância na identificação e confirmação de variáveis importantes para realização de estudos futuros. Como afirmam Liberato et al. (1999) e Barbosa et al. (2005), os pesquisadores tendem a avaliar o maior número de características; no entanto, é possível que parte dessas contribuam pouco ou nada para a caracterização, não agregando melhorias na precisão, além de tornar mais complexa a análise e interpretação dos dados.

As variáveis presentes no grupo 2 são relacionadas ao consumo, como está apresentado na Tabela 2. Diferente do direcionamento tomado no primeiro grupo, nesse, em questão, praticamente todas as variáveis do grupo contribuíram para que outras fossem agregadas. Segundo Hair (2009), a importância das variáveis para a formação dos grupos é observada mediante o distanciamento que essas têm do centro destes. Conforme observado, a maior parte das distâncias em relação ao centro é pequena, confirmando a importância de cada uma para formação do grupo.

As variáveis presentes no grupo 3, Tabela 2, são relacionadas às medidas morfométricas; no entanto, como é possível perceber, todas apresentam maior distância do centro do grupo. Essa tendência pode indicar que talvez exista a possibilidade de o conjunto de características não está representado da melhor forma. Uma solução seria realizar um novo rearranjo das variáveis, visando a melhor distribuição das características, ação que colocaria à prova a possibilidade de alguma variável, presente no conjunto de dados, ter maior poder de atração para formação de um novo grupo ou de um já existente, fato que poderia ser representado pelo menor distanciamento do centro do grupo.

No grupo 4, as características com maior influência para sua formação foram: RCF (Rendimento de Carcaça Fria) e RCQ (Rendimento de Carcaça Quente), como mostra a Tabela 2, ambas, presentes no segundo componente principal, avaliado na análise de componentes principais. A influência dessas duas características relacionadas ao rendimento da carcaça é evidenciada no estudo realizado por Guedes (2017), quando recomenda o uso de ambas em experimentos futuros.

A distribuição dos grupos aconteceu como previsto, uma vez que na análise de k-médias é indicado o número de grupos, antes de serem formados, geralmente baseados na hipótese pré-determinada de confrontar as principais variáveis de formação

dos grupos com as obtidas na análise de componentes principais nas quatro primeiras combinações. Como informado, na distribuição obtida na análise de componentes principais, o direcionamento das variáveis indicou a possibilidade de formar grupos relacionados a medidas morfométricas, bem como de desempenho animal. Almeida et al. (2015) ressaltaram a importância que as medidas morfométricas e os índices zootécnicos de desempenho representam para a caracterização fenotípica de grupos raciais, indispensáveis ao melhoramento animal, ao conhecimento da conformação adequada dos animais e ao estabelecimento da relação conformação e funcionalidade.

As maiores distâncias representam as características que menos contribuíram para formar o grupo. Fazendo uma comparação com as variáveis que menos contribuíram para explicar a variação total dos dados, na análise de componentes principais, considerando os CP's (componentes principais) selecionados, é possível eleger algumas que possivelmente são características que pouco representam o resultado obtido e, talvez, sejam redundantes, sendo elas: AOL, CEE, CFDN e Profundidade de Tórax.

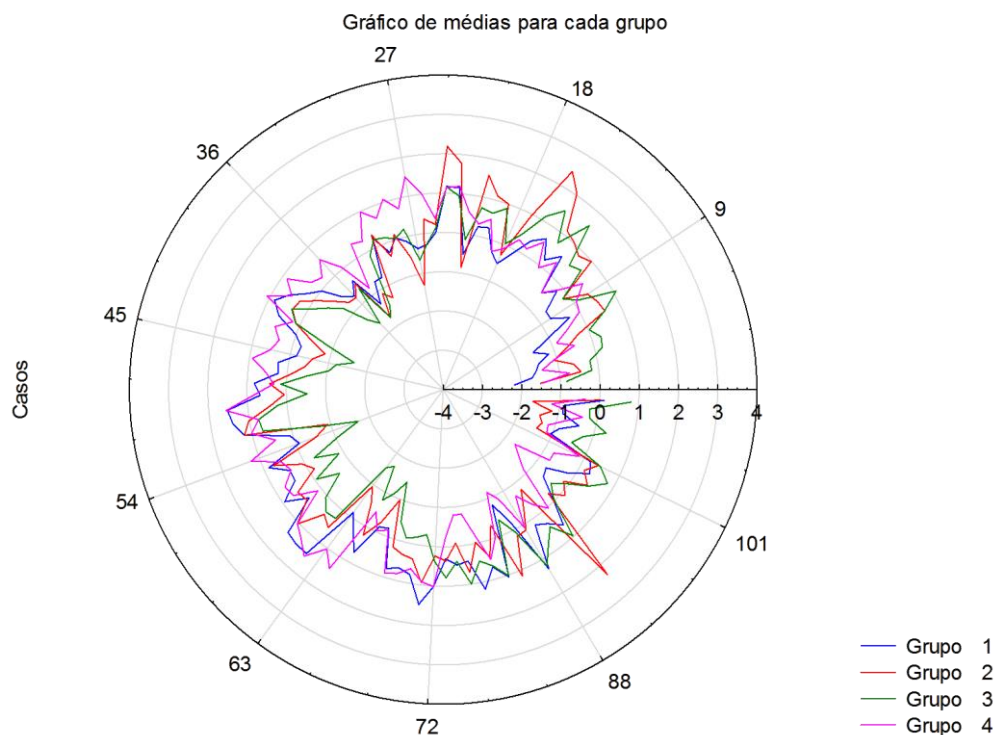


Figura 2. Gráfico de médias por grupo, obtido pela análise de agrupamento K-médias.

Variáveis com alta capacidade de atrair novos componentes para o grupo apresentam menor distância do centro, estando representado na Figura 2 o número de

casos usados para cada grupo. Conforme análise estatística, percebe-se a maior homogeneidade do grupo 3, visto que a maior parte das características contribuiu para formar o grupo. Ao mesmo passo que é possível perceber a maior oscilação dos casos presentes no grupo 2, cenário que reflete na composição do grupo e conseqüentemente nas distâncias apresentadas.

4. Conclusão

A análise de agrupamento permitiu identificar as variáveis mais influentes na formação dos grupos pré-determinados. Os resultados encontrados com a aplicação da análise de Cluster agregaram maior confiabilidade aos obtidos com a aplicação de outra técnica de análise multivariada, a de componentes principais. As características de maior influência foram: PCF, PCQ, PCVZ, RCF e RCQ, essas relacionadas à carcaça. Já as de desempenho foram: CMS, CMO, CPB, podendo essas ser utilizadas em estudos relacionados ao melhor rendimento de carcaça e obtenção de melhores resultados na escolha da dieta de animais, além de economizar tempo, uma vez que o direcionamento pode ser dado as variáveis com maior poder de atração.

5. Referências Bibliográficas

ALMAS, S. P. Análise das estatísticas relacionadas ao jogo que discriminam as equipes vencedoras das perdedoras no basquetebol profissional brasileiro. **Revista Brasileira de Educação Física e Esporte**. 29 (4), p. 551- 558, São Paulo, 2015.

ALMEIDA, J. C. S. et al. Desempenho, medidas corporais, rendimentos de carcaça e cortes, e qualidade de carne em cordeiros alimentados com resíduos da agroindústria processadora de frutas. **Semina: Ciências Agrárias**, Londrina, v.36, n.1, p.541-556, jan/fev, 2015.

BARBOSA, L.T. et al. Seleção de variáveis de desempenho de suínos por meio da análise de componentes principais. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.57, n.6, p. 805-810, 2005b.

COSTA, R. M. **Uso de informação \genômica para estimação de parâmetros genéticos para caracterização de crescimento e carcaça em bovinos Nelore**. Dissertação (Mestrado em Genética e Melhoramento Animal). Faculdade de Ciências Agrárias e Veterinárias – Unesp. 61p. 2019.

CRUZ, D. A. C. et al. Cluster analysis of breeding values for milk yield and lactation persistency in Guzera cattle. **Ciência Rural**, 46. 2016.

- GORDON, A. D. **Classification**. Chapman and Hall Ed., 1981.
- GUEDES, D. G. P. et al. **Técnicas estatísticas multivariadas aplicadas à caracterização de carcaça de ovinos da raça Morada Nova**. Recife, PE: UFRPE, 2017, 99P. Tese (Doutorado Integrado em Zootecnia) - Universidade Federal Rural de Pernambuco, Universidade Federal da Paraíba e Universidade Federal do Ceará, 2017.
- HAIR, J. F. JR.; ANDERSON, R. E. et al. **Multivariate Data Analysis**. Prentice Hall. 1995.
- HAIR, J.R. et al. **Multivariate Data Analysis**. 6ª edição. Upper Saddle River, NJ: Pearson Prentice Hall, 2009.
- LINDEN, R. Técnicas de Agrupamento. **Revista Sistema de Informação da FSMA**. N. 4, p.18- 36, 2009.
- LATTIN, J. M. DOUGLAS, C.; PAUL, E. G. **Análise de dados multivariados**. Cengage Learning, 455 p. 2011.
- LIBERATO, J.R.; VALE, F.X.R.; CRUZ, C.D. Técnicas estatísticas de análise multivariada e a necessidade de o fitopatologista conhecê-las. **Fitopatologia Brasileira**, v.24, p.5-8, 1999.
- MALHORTA, N. **Pesquisa de marketing: uma orientação aplicada**. Trad. Laura Bocco. Porto Alegre: Bookman, ed.4. 2006.
- MORAES, M. B. C. **Análise multivariada aplicada à contabilidade**. Universidade de São Paulo, Disciplinas online, 2016.
- NETO, M.R. T. et al. "Descrição do crescimento de ovinos Santa Inês utilizando modelos não-lineares selecionados por análise multivariada". **Revista Brasileira de Saúde e Produção Animal** 17.1 (2016).
- REGAZZI, A. J. **Análise multivariada**. Universidade Federal de Viçosa. Centro de Ciências Exatas e Tecnológicas. Departamento de Informática, 116p. 2001.
- SEIDEL E. J. et al. Comparação entre o método Ward e o método K-médias no agrupamento de produtores de leite. **Ciência e Natura** 30:7- 15, 2008.

Considerações Finais

A escolha de uma técnica para análise de dados que consiga atingir o objetivo pretendido no estudo é extremamente importante, principalmente quando se busca avaliar as possíveis relações entre as variáveis em teste. A análise multivariada corresponde à interpretação de um grande número de informações, de modo que seja possível utilizar simultaneamente todas as variáveis na interpretação do conjunto de dados.

As técnicas de análise multivariada possibilitam a integração das correlações existentes à análise das variáveis, de modo que a interpretação é feita baseada na influência que cada variável tem na formação do resultado obtido. Este fato contribui consideravelmente para a seleção das variáveis que possuem melhor capacidade de explicar a variação total dos dados. As técnicas multivariadas aplicadas em conjunto possibilitam melhores respostas, visto que, em sua grande maioria, são complementares.

A aplicação das técnicas de análise de componentes principais e análise de Cluster (agrupamento) possibilitaram, de forma eficiente, a identificação das variáveis de maior influência para a caracterização da carcaça de ovinos da raça Santa Inês, além de reduzir a dimensionalidade dos dados originais e indicar as variáveis que têm maior influência para a formação das combinações lineares.