

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

MARILENE DOS SANTOS MACIEL

**AVALIAÇÃO DE CARCAÇAS OVINAS DA RAÇA SANTA INÊS: UMA
ABORDAGEM MULTIVARIADA**

**RECIFE
2022**

MARILENE DOS SANTOS MACIEL

**AVALIAÇÃO DE CARCAÇAS OVINAS DA RAÇA SANTA INÊS: UMA
ABORDAGEM MULTIVARIADA**

Tese apresentada ao Programa de Pós-Graduação em Zootecnia da Universidade Federal de Pernambuco para obtenção do título de Doutor em Zootecnia

Área de concentração: Zootecnia

Orientador (a): Prof^a. Dr^a. Maria Norma Ribeiro

Coorientador(es): Prof^o. Dr. George Rodrigo Beltrão da Cruz

Prof^o. Dr. Francisco Fernando Ramos de Carvalho

**RECIFE
2022**

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- M152a Maciel, Marilende dos Santos
AVALIAÇÃO DE CARCAÇAS OVINAS DA RAÇA SANTA INÊS: UMA ABORDAGEM MULTIVARIADA /
Marilende dos Santos Maciel. - 2022.
94 f. : il.
- Orientadora: Maria Norma Ribeiro.
Coorientadora: George Rodrigo Beltrao da Cruz e Francisco Fernando Ramos de Carvalho.
Inclui referências.
- Tese (Doutorado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Zootecnia, Recife,
2022.
1. Classificação de carcaça. 2. Escores fatoriais . 3. Rotações ortogonais . 4. Medidas morfométricas . 5. Seleção
stepwise. I. Ribeiro, Maria Norma, orient. II. Cruz, George Rodrigo Beltrao da, coorient. III. Carvalho, Francisco
Fernando Ramos de, coorient.



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

**AVALIAÇÃO DE CARÇAÇAS OVINAS DA RAÇA SANTA INÊS: UMA
ABORDAGEM MULTIVARIADA**

Tese elaborada por:

MARILENE DOS SANTOS MACIEL

Aprovado em/...../.....

BANCA EXAMINADORA

Prof^a. Dr^a. Maria Norma Ribeiro
Universidade Federal Rural de Pernambuco (UFRPE)

Prof. Dr. Roberto Germano Costa
Universidade Federal da Paraíba (UFPB)

Prof^a. Dr^a. Maria Antônia Revidatti
Universidad Nacional del Nordeste (UNNE)

Prof^a. Dr^a. Janaína Kelli Gomes Arandas
Faculdade Novo Horizonte (FNH)

Prof^a. Dr^a. Laura Leandro da Rocha
Universidade Federal Rural do Pernambuco (UFRPE)

*Ao meu pai, **Felisberto Ferreira Maciel**, e à minha mãe, **Mariene Evangelista dos Santos Maciel**. De vocês recebi o dom mais precioso do universo: a vida. Por isso, já seria eternamente grata, mas além dela, vocês ainda me deram amor, carinho, conforto e cultivaram em mim, ainda criança, todos os valores que me transformaram em um adulto responsável e consciente. Também abriram as portas para meu futuro, muitas vezes sacrificando os seus sonhos em favor dos meus. Mesmo nas horas que meus sonhos pareciam distantes e inatingíveis e em que o estudo parecia estar além dos nossos limites, vocês estavam presentes, à disposição para me ouvir, orientar, encorajar e fazer acreditar que somos vencedores em potencial. Tantas vezes meu cansaço e minha preocupação foram sentidos e compartilhados por vocês, num colo, num afago ou mesmo num silêncio. Por isso, existe em mim uma imensa vontade de alcançar o tão sonhado objetivo e ver o sorriso de satisfação e orgulhos estampado em seus rostos. Vocês são meus verdadeiros mestres e doutores da vida. Amo vocês!*

DEDICO

AGRADECIMENTOS

A Deus, por ser essencial em minha vida, autor do meu destino, meu guia, socorro presente na hora da angústia;

À Universidade Federal Rural de Pernambuco (UFRPE) e ao Programa de Pós-Graduação em Zootecnia/UFRPE, pela oportunidade de realização deste trabalho;

À Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES), pela concessão de bolsa;

À minha querida orientadora, Prof^a. Dr^a. Maria Norma Ribeiro. Agradeço por fazer conhecer o lado humano da carreira acadêmica, ao qual aprendi a admirar e respeitar, pela paciência que teve nos momentos mais difíceis, pelo incentivo e, principalmente, por acreditar em minha capacidade.

Aos meus coorientadores, Prof. Dr. George Rodrigo Beltrão da Cruz e Prof. Dr. Francisco Fernando Ramos de Carvalho, por prestarem toda a assistência necessária nesta jornada acadêmica;

A todos os professores do Programa de Pós-Graduação em Zootecnia da UFRPE/Recife, pela transmissão de conhecimento por meio das disciplinas cursadas.

À Dr^a. Janaína Kelli Gomes Arandas, exemplo de pesquisadora e amiga. Agradeço por todos os seus ensinamentos, paciência e confiança ao longo das supervisões das minhas atividades e incentivos que tornaram possível a conclusão deste trabalho. Sou sua fã;

Ao Grupo de Conservação de Recursos Genéticos Animal (CRGAN), que exerceu uma base extra para minha formação acadêmica. As conversas e debates durante as reuniões foram riquíssimas;

Ao meu querido colega, amigo e irmão Paulo Otávio. Durante todo o período do doutorado, você foi meu equilíbrio; agradeço por sua paciência e incentivo. Te admiro muito, irmãozinho;

Aos meus queridos irmãos Antusa, Alzemira, Ailson, Geilson e Maisa. Vocês são extraordinários, apesar de todas nossas brigas, apesar de todos os erros, vocês são minha única certeza de que fariam coisas por mim sem pedir nada em troca. Obrigada por tudo. Amo vocês;

Aos meus queridos sobrinhos Izadora, Andrei, Karoline e Henrique. Vocês proporcionaram o lado mais leve da vida, a inocência mais doce e singela, os sorrisos mais espontâneos e

principalmente de fazer lembrar a importância de não deixarmos de sermos crianças mesmo sendo adultos. Meus amores;

À Chrislanne e Jessica Berly, minha família no Recife durante esta jornada. Agradeço por compartilharem comigo tanto a vida acadêmica como pessoal;

A todos meus amigos e colegas da Ruralinda, em especial Taile Katiele, Zaqueu, Tomás Guilherme, Jasiel e Leonardo;

E a todos que, de forma direta ou indireta, contribuíram para a realização desta conquista, pelas alegrias, tristezas e dores compartilhadas com vocês, as pausas entre um parágrafo e outro, melhoraram tudo que tenho produzido na vida.

Meu muito obrigada!

“Todos somos pessoas comuns caminhando pela rua que Deus pôs a nossa frente. Não importa se você é rico ou pobre, ou se está em um lugar entre os dois. Todos somos desabrigados. Cada um de nós nesse mundo. Apenas tentando voltar para casa”.

(Denver Moore)

1 **AValiação de Carcaças Ovinas da Raça Santa Inês: Uma** 2 **Abordagem Multivariada**

3
4 **RESUMO** – A análise multivariada corresponde a um conjunto de métodos que utilizam
5 simultaneamente as informações de todas as variáveis respostas na interpretação do conjunto de dados,
6 levando em consideração as correlações existentes entre elas. Dentre os conjuntos de métodos
7 multivariados podemos destacar a análise fatorial, a análise de regressão múltipla e a análise
8 discriminante multivariada. Esses métodos têm em comum o objetivo de reduzir a dimensionalidade dos
9 dados. Considerando que na avaliação de carcaça um número elevado de variáveis é utilizado, existe a
10 possibilidade que muitas delas contribuem pouco para a discriminação dos indivíduos avaliados, por
11 serem redundantes em virtude das altas correlações. Com base no exposto, objetivou-se, por meio da
12 análise fatorial, avaliar a adequação desta e dos métodos rotacionais; estabelecer um modelo
13 multivariado utilizando duas técnicas estatísticas multivariadas complementares, Análise Fatorial e
14 Regressão Múltipla *Stepwise*, para prever a composição dos tecidos (músculo, osso e gordura); e
15 distinguir grupos de ovinos em função do escore de conformação e acabamento de carcaça, e identificar
16 as variáveis que mais contribuem na diferenciação, mediante análise discriminante multivariada. Foram
17 utilizadas informações de 122 ovinos da raça Santa Inês, compreendendo 24 características de carcaças:
18 peso de corpo vazio (PCV); peso vivo ao abate (PVA); peso de carcaça quente (PCQ); peso de carcaça
19 fria (PCF); rendimento biológico (RB); rendimento de carcaça quente (RCQ); rendimento de carcaça
20 fria (RCF); comprimento externo da carcaça (CEC); comprimento interno da carcaça (CIC);
21 comprimento da perna (CP); largura da garupa (LG); largura do tórax (LT); perímetro da garupa (PG);
22 perímetro do pernil (PP); perímetro do tórax (PT); índice de compacidade da perna (ICC); pernil (PER);
23 lombo (LOM); costela (COS); pescoço (PES); paleta (PAL); rendimento do pernil (RPER); rendimento
24 do lombo (RLOM); rendimento da costela (RCOS); rendimento do pescoço (RPES); rendimento da
25 paleta (RPAL). O teste de esfericidade de Bartlett e KMO (Kaiser-Meyer-Olkin) indicaram que as
26 variáveis de carcaças avaliadas foram adequadas para aplicação da análise fatorial. Dentre as rotações
27 testadas, a rotação ortogonal Varimax foi a que apresentou estrutura mais simples para interpretação dos
28 fatores construídos. A utilização das variáveis latentes da análise fatorial em modelos de regressão
29 múltipla elimina o problema de multicolinearidade das variáveis explicativas, melhorando, assim, a
30 acurácia da interpretação dos resultados por propor um melhor ajuste do modelo matemático. Porém, os
31 valores dos coeficientes de determinação (R^2) foram moderados para proporção muscular e gordura total
32 e baixo para a proporção óssea, indicando que variáveis independentes mais adequadas devem ser usadas
33 para melhor prever a proporção dos tecidos em ovinos da raça Santa Inês. Na análise discriminante
34 multivariada, as variáveis com maior poder discriminatório para os escores de conformação de carcaça
35 foram PCF, CEC e PES, e para acabamento de carcaça foram PVA, CEC e PT. A análise discriminante
36 multivariada mostrou-se eficiente para alocar os animais em seus grupos de origem (escores de carcaça).

37
38 **PALAVRAS-CHAVE:** Classificação de carcaça. Escores fatoriais. Rotações ortogonais. Medidas
39 morfométricas. Seleção *stepwise*.

EVALUATION OF SANTA INES SHEEP CARCASSES: A MULTIVARIATE APPROACH

ABSTRACT – Multivariate analysis corresponds to a set of methods that simultaneously use information from all response variables in the interpretation of the data set, taking into account the correlations between them. Among the sets of multivariate methods we can highlight factor analysis, multiple regression analysis and multivariate discriminant analysis. These methods have in common the objective to reduce the dimensionality of the data. Considering that a large number of variables are used in the carcass evaluation, there is the possibility that many of them contribute little to the discrimination of the evaluated individuals, as they are redundant due to the high correlations. Based on the above, the objective was, through factor analysis, to evaluate the adequacy of factor analysis and rotational methods; establish a multivariate model using two complementary multivariate statistical techniques, Factor Analysis and Stepwise Multiple Regression to predict tissue composition (muscle, bone and fat) and; distinguish groups of sheep as a function of conformation score and carcass finish and identify the variables that most contribute to differentiation, through multivariate discriminant analysis. Information from 122 sheep of the Santa Inês breed was used, comprising 24 carcass characteristics: Empty body weight (EBW); Body weight at slaughter (BWS); Hot carcass weight (HCW); Cold carcass weight (CCW); Biological yield (BY); Hot carcass yield (HCY); Cold carcass yield (CCY); External carcass length (EXL); Internal carcass length (ICL); Rump width (RW); Chest width (CW); Rump perimeter (RP); Leg perimeter (LP); Chest perimeter (CP); Leg (LEG); Loin (LOI); Ribs (RIB); Neck (NEC); Blade Shoulder (BLA); Leg yield (LEGY); Loin yield (LOIY); Rib yield (RIBY); Neck yield (NECY); Blade Shoulder yield (BLAY). The sphericity test of Bartlett's and KMO (Kaiser-Meyer-Olkin) indicated that the carcass variables evaluated were adequate for the application of factor analysis. Among the rotations tested, the Varimax orthogonal rotation presented the simplest structure for interpreting the constructed factors. The use of latent variables from factor analysis in multiple regression models eliminates the problem of multicollinearity of the explanatory variables, thus improving the accuracy of the interpretation of results by proposing a better adjustment of the mathematical model. However, the values of the coefficients of determination (R^2) were moderate for the proportion of muscle and total fat and low for the proportion of bone, indicating that more adequate independent variables should be used to better predict the proportion of tissues in Santa Inês sheep. In the multivariate discriminant analysis, the variables with the highest discriminatory power for carcass conformation scores were CCW, EXL and NEC and for carcass finishing were BWS, EXL and CP. Multivariate discriminant analysis proved to be efficient to allocate the animals to their original groups (carcass scores).

KEYWORDS: Carcass classification. Factorial scores. Orthogonal rotations. Body measurements. *Stepwise* selection.

LISTA DE ILUSTRAÇÕES

CAPÍTULO I

Figura 1 – Comparação de vizinhos próximos (k) avaliados em experimentos com carne suína e bovina	27
--	----

LISTA DE TABELAS

CAPÍTULO I

Tabela 1 – Diretrizes para identificação de cargas fatoriais significantes com base em tamanho de amostra.....	21
Tabela 2 – Diretrizes para interpretar os valores de VIF (Fator de inflação da variância).....	24

CAPÍTULO II

Tabela 1 – Média (μ) e desvio padrão (DP) das variáveis de carcaças ovinas da raça Santa Inês.....	35
Tabela 2 – Critério de Kaiser-Meyer-Olkin (KMO).....	36
Tabela 3 – Diretrizes para identificação de cargas fatoriais significantes com base em tamanho de amostra.....	37
Tabela 4 – Correlação linear simples (abaixo da linha diagonal) e nível de probabilidade de significância (acima da linha diagonal) entre características de carcaças de ovinos da raça Santa Inês.....	40
Tabela 5 – Adequação da amostragem pelo teste de esfericidade de Barlett e o teste de KMO (Kaiser-Meyer-Olkin)	41
Tabela 6 – Teste de normalidade multivariada Mardia para as variáveis de carcaças ovinas da raça Santa Inês.....	42
Tabela 7 – Autovalores da matriz de correlação das 24 variáveis analisadas pelo critério de Kaiser.....	43
Tabela 8 – Cargas fatoriais dos seis fatores, sem a utilização de rotação dos fatores e suas respectivas comunalidades.....	44
Tabela 9 – Correlações entre os fatores pela rotação oblíqua Oblimin direta.....	45
Tabela 10 – Cargas fatoriais dos seis fatores, obtido pela rotação oblíqua Oblimin direta e suas respectivas comunalidades.....	46
Tabela 11 – Correlações entre os fatores pelo método de rotação oblíqua Promax.....	47
Tabela 12 – Cargas fatoriais dos seis fatores obtidas pela rotação oblíqua Promax e suas respectivas comunalidades.....	48
Tabela 13 – Cargas fatoriais dos seis fatores retidos pelo método de Kaiser, obtido pela rotação ortogonal Quartmax e suas respectivas comunalidades.....	50
Tabela 14 – Cargas fatoriais dos seis fatores obtidas pela rotação ortogonal Equamax e suas respectivas comunalidades.....	52
Tabela 15 – Cargas fatoriais dos seis fatores obtidas pela rotação ortogonal Varimax e suas respectivas comunalidades.....	54

CAPÍTULO III

Tabela 1 – Principais informações dos experimentos.....	61
Tabela 2 – Estatísticas descritivas para variáveis dependentes.....	65
Tabela 3 – Estatísticas descritivas para variáveis independentes.....	66
Tabela 4 – Correlação de Pearson (r) entre a composição dos tecidos e variáveis de carcaça.....	67
Tabela 5 – Equações de predição para a composição dos tecidos de acordo a análise de regressão múltipla <i>stepwise</i> das variáveis originais.....	70
Tabela 6 – Adequação da amostragem pelo teste de esfericidade de Barlett e o teste de KMO (Kaiser-Meyer-Olkin)	72
Tabela 7 – Resultados da análise fatorial aplicada às variáveis independentes.....	73

Tabela 8 – Equações de predição para a composição dos tecidos da análise de regressão múltipla <i>stepwise</i> das variáveis latentes.....	75
---	----

CAPÍTULO IV

Tabela 1 – Estatísticas descritivas da análise discriminante para os grupos de escores de conformação (2 e 3) e acabamento (2 e 3)	82
Tabela 2 – Teste de normalidade multivariada Mardia para as variáveis independentes das carcaças ovinas.....	83
Tabela 3 – Variáveis independentes selecionadas pelo procedimento <i>stepwise</i> para os escores de conformação (2 e 3) de carcaças ovinas e a D^2 de Mahalanobis.....	84
Tabela 4 – Coeficientes canônicos padronizados para a variável canônica (CAN1) para os escores de conformação (2 e 3)	86T
Tabela 5 – Número de observações e percentual de classificação dos grupos de escores de conformação (2 e 3)	87
Tabela 6 – Variáveis independentes selecionadas pelo procedimento <i>stepwise</i> , para os escores de acabamento (2 e 3) de carcaças ovinas e a D^2 de Mahalanobis.....	87
Tabela 7 – Coeficientes canônicos padronizados para a variável canônica (CAN1) para os escores de acabamento (2 e 3)	88
Tabela 8 – Número de observações e percentual de classificação dos grupos de escores de acabamento (2 e 3)	89

LISTA DE ABREVIATURAS E SIGLAS

μ - média
C – comunalidade
CEC – comprimento externo da carcaça
CIC – comprimento interno da carcaça
cm – centímetro
COS – costela
CP – comprimento da perna
DP – desvio padrão
HSI – Hue Saturation Intensity
ICC – índice de compactidade da carcaça
kg – quilograma
KMO – Kaiser-Meyer-Olkin
K-NN – vizinho mais próximo
LDA – análise discriminante linear
LG – largura da garupa
LOM – lombo
MSE – erro quadrado médio
 n – número
 p – variáveis
PAL – paleta
PCF – peso de carcaça fria
PCQ – peso de carcaça quente
PCV – peso de corpo vazio
PER – pernil
PES – pescoço
PP – perímetro do pernil
PR – perda por resfriamento
PT – perímetro do tórax
PVA – peso vivo ao abate
QDA – análise discriminante quadrática
 R^2 – coeficiente de determinação
RAG – resíduo agroindustrial de goiaba
RB – rendimento biológico
RCF – rendimento de carcaça fria
RCOS – rendimento de costela
RCQ – rendimento de carcaça quente
RLOM – rendimento de lombo
RPAL – rendimento de paleta
RPER – rendimento de pernil
RPES – rendimento de pescoço
SRD – sem raça definida
TGI – trato gastrointestinal

SUMÁRIO

RESUMO.....	vii
ABSTRACT.....	viii
1. INTRODUÇÃO.....	14
CAPÍTULO I – Revisão de literatura	16
1. ANÁLISE FATORIAL.....	17
2. ADEQUAÇÃO DE DADOS PARA A ANÁLISE FATORIAL.....	18
3. USO DOS ESCORES FATORIAIS DA ANÁLISE FATORIAL NA ANÁLISE DE REGRESSÃO MÚLTIPLA.....	21
4. ANÁLISE DISCRIMINANTE MULTIVARIADA.....	24
4.1 Vizinho mais próximo (K-NN)	26
5. REFERÊNCIAS.....	28
CAPÍTULO II – Adequação da análise fatorial e dos métodos de rotação para a caracterização de carcaças de ovinos da raça Santa Inês	32
RESUMO.....	33
ABSTRACT.....	33
1. INTRODUÇÃO.....	33
2. MATERIAL E MÉTODOS.....	34
3. RESULTADOS E DISCUSSÃO.....	38
4. CONCLUSÃO.....	56
5. REFERÊNCIAS.....	56
CAPÍTULO III – Comparação análise fatorial e regressão múltipla para predição da composição dos tecidos da carcaça de ovinos da raça Santa Inês	58
RESUMO.....	59
ABSTRACT.....	59
1. INTRODUÇÃO.....	59
2. MATERIAL E MÉTODOS.....	60
3. RESULTADOS E DISCUSSÃO.....	64
4. CONCLUSÃO.....	76
5. REFERÊNCIAS.....	77
CAPÍTULO IV – Diferenciação de ovinos, de acordo com escores de conformação e acabamento de carcaça: um estudo usando análise discriminante canônica	79
RESUMO.....	80
ABSTRACT.....	80
1. INTRODUÇÃO.....	80
2. MATERIAL E MÉTODOS.....	81
3. RESULTADOS E DISCUSSÃO.....	84
4. CONCLUSÃO.....	89
5. REFERÊNCIAS.....	90

1. INTRODUÇÃO

As características da carcaça são essenciais em sistemas de produção de ovinos de corte, pois representam a matéria-prima fundamental para as indústrias que processam produtos de origem animal. Planejar, projetar, executar e controlar processos que acercam a avaliação de características de carcaça costuma ser complexo. Muitos dados e informações são gerados, e transformá-los em conhecimento requer avaliações mais confiáveis. Para tais sistemas de avaliações complexas, as pesquisas atuais têm sido conduzidas usando métodos multivariados, como a análise de componentes principais (ARSALANE *et al.*, 2017; HOSSEINPOUR; ILKHCHI; AGHBASHLO, 2019; KAMRUZZAMAN *et al.*, 2012, 2013; PAPADOPOULOU *et al.*, 2013), análise fatorial (ÇELIK *et al.*, 2018; DASKIRAN; KESKIN; BINGOL, 2017; ÖNK; SARI; GÜRCAN, 2018; TAHTALI, 2019), análise discriminante (CLUFF *et al.*, 2013; KAMRUZZAMAN *et al.*, 2012; NAGANATHAN *et al.*, 2015; ZAPOTOCZNY; SZCZYPIŃSKI; DASZKIEWICZ, 2016), dentre outras. Diversos processos de avaliações apresentam natureza multivariada, uma vez que algumas características só podem ser completamente caracterizadas pela avaliação de características de qualidade distintas (MARQUES *et al.*, 2020).

Nesse sentido, as estatísticas multivariadas oferecem um conjunto de ferramentas eficientes para analisar padrões de correlações complexas entre as variáveis, além de reduzir a redundância entre elas. Para que ocorra a redução na dimensionalidade dos dados, uma análise exploratória dos dados torna-se necessário. Como descrito por Arsalane *et al.* (2018), depois que a aquisição de dados é concluída e as variáveis calculadas, geralmente alguma forma de redução de dimensionalidade é muito importante e necessário, pois permite coletar menos dados e gera um modelo mais simples e de fácil interpretação.

A análise fatorial multivariada, por meio da criação de novas variáveis (fatores ou variáveis latentes), retém o máximo de informações das variáveis originais. O surgimento dessas variáveis latentes ocorre da combinação linear das variáveis originais, em que os escores fatoriais representam as somas ponderadas das variáveis originais com perdas mínimas de informações úteis, sendo que os primeiros fatores retêm a maior parte da variação presente em todas as variáveis originais. Esses fatores podem ser interpretados independentemente, o que permite uma visão geral da estrutura de dados. No entanto, assim como ocorre com outras técnicas multivariadas, a execução da análise fatorial requer tomada de decisões, dentre elas, a escolha dos métodos de rotações, que podem ser ortogonais ou oblíquos, dentre os quais devemos optar pela rotação ortogonal, quando os fatores não são correlacionados, e pelas rotações oblíquas para fatores correlacionados.

Por outro lado, em situações em que o objetivo é determinar a força da relação entre uma variável dependente e um conjunto de variáveis independentes, que é o caso da análise de regressão múltipla, pode haver variáveis independentes que pouco influenciam a variável dependente. Nesse contexto, a seleção de um subconjunto de covariáveis que maximiza a capacidade de previsão é necessária. Geralmente, o procedimento *stepwise* é usado para selecionar aquelas variáveis que mais influenciam a variável dependente, podendo, assim, diminuir o número de variáveis a compor a equação do modelo (ALVES; LOTUFO; LOPES, 2013). Entretanto, a utilização desse procedimento apresenta uma desvantagem, que é a possibilidade de multicolinearidade. A multicolinearidade significa que têm variáveis independentes que são altamente correlacionadas umas com as outras. Como salientado por Ali *et al.* (2015), se houver correlações altas ($>0,80$) entre as variáveis independentes, as análises de regressão múltipla podem produzir estimativas enviesadas devido ao problema de multicolinearidade. Para abordar essa limitação, a análise fatorial multivariada, por meio dos escores fatoriais pode ser adequada e usada para previsão da análise de regressão múltipla, resolvendo, assim, o problema da multicolinearidade. Outro ponto de destaque, em geral, é que na experimentação animal, as variáveis estudadas são definidas em grupos *a priori*; a exemplo disso, podemos citar as características de carcaças ovinas que são obtidas em amostras dentro de cada grupo. Considerando que geralmente os pesquisadores estão interessados em identificar a contribuição de cada variável na separação dos vários grupos, a análise discriminante multivariada pode nortear para um método complementar na análise dos dados.

A análise discriminante multivariada é empregada para identificar a contribuição de cada variável escolhida na separação de grupos, encontrando uma combinação linear de k variáveis medidas (ALKARKHI; ALQARAGHULI, 2020). Em outras palavras, a análise discriminante é usada para selecionar variáveis contínuas que discriminam entre dois ou mais grupos e determinar combinações lineares das variáveis, ou seja, variáveis que fornecem a discriminação máxima entre grupos (JEON *et al.*, 2013), além de ser útil para prever a qual grupo uma observação pertence, com base no conhecimento das variáveis quantitativas em um conjunto de combinações lineares dessas variáveis (PARK *et al.*, 2002).

Assim, este trabalho teve por objetivo avaliar a aplicação de algumas das principais técnicas de análise multivariada em um conjunto de variáveis referentes a características de carcaça de ovinos da raça Santa Inês, a fim de verificar a adequação da análise fatorial e dos métodos rotacionais, buscar uma alternativa para solucionar o problema da multicolinearidade na análise de regressão múltipla, estudar a associação entre grupo de variáveis e selecionar aquelas variáveis mais importantes e com maior poder discriminatório.

CAPÍTULO I

Revisão de literatura

1. ANÁLISE FATORIAL

A análise fatorial é um ramo da estatística multivariada que foi inicialmente desenvolvido por psicólogos, para lidar principalmente com hipóteses sobre a organização das habilidades mentais, considerando a estrutura de correlação das variáveis de testes cognitivos (LAWLEY; MAXWELL, 1962). O objetivo da análise fatorial é descrever a estrutura de covariância entre variáveis, reduzindo dimensões subjacentes existentes entre elas. Este procedimento é baseado no fato de que qualquer conjunto de variáveis contínuas observado em cada uma das n unidades experimentais pode ser transformado em um novo conjunto de variáveis (GOMES *et al.*, 2013). Essas novas variáveis são chamadas de fatores, e cada variável é uma combinação linear das variáveis originais. Os fatores também são conhecidos com variáveis latentes, ou seja, geram relações não observáveis, conseqüentemente, a redundância entre as variáveis pode ser reduzida, facilitando a análise.

Na análise fatorial, cada variável é descrita como uma função linear em termos de fatores comuns e fatores específicos (MARQUES *et al.*, 2020). Os fatores comuns explicam as variâncias e covariâncias das variáveis originais em uma variância comum (compartilhada com algumas ou todas as variáveis) e os fatores específicos explicam apenas as variâncias das variáveis originais (específica para uma determinada variável) (GOMES *et al.*, 2013).

A análise fatorial ainda é amplamente aplicada como método de análise multivariada no campo da Psicologia, mas os avanços recentes ajudaram disseminar o método para outros campos, com Ciências Sociais (VICTOR *et al.*, 2018), Biologia (BUETTNER *et al.*, 2017), Economia (HOQUE *et al.*, 2018) e Engenharia (PATRIARCA *et al.*, 2018).

Na produção animal, a análise fatorial vem sendo aplicada para avaliar variáveis relacionadas à produção, composição e propriedades de coagulação do leite (MACCIOTTA *et al.*, 2004, 2012; MANCA *et al.*, 2016; TODARO; SCATASSA; GIACCONE, 2005) e perfil de ácidos graxos no leite (CONTE *et al.*, 2016; CORREDDU *et al.*, 2017; MELE *et al.*, 2016; TURINI *et al.*, 2020) em animais destinados à produção de leite. Em todos os estudos supracitados, a análise fatorial mostrou-se adequada para estudar e interpretar sistemas multivariados complexos por meio da extração de poucas variáveis (variáveis latentes), com claro significado técnico e biológico.

Em animais destinados à produção de carne, um estudo foi realizado para analisar as inter-relações entre pesos e medidas de carcaça em caprinos (GOMES *et al.*, 2013). A aplicação da análise fatorial permitiu que em um conjunto de 32 variáveis originais fossem condensadas em quatro fatores comuns, que explicaram 77,2% da variabilidade nas 32 variáveis originais, restando 22,8% para os fatores específicos, característicos de cada variável. As variáveis que

continham as maiores partes das informações relevantes do grupo original de variáveis foram peso de carcaça quente, perímetro da garupa, comprimento da perna e medidas do tecido da 3ª e 4ª vértebra. Em outro estudo, Cadavez (2021), avaliando características de carcaças ovinas, a análise fatorial também mostrou-se eficiente para entender as inter-relações entre as variáveis em um grande conjunto de dados. O conjunto de dezoito variáveis originais pôde ser reduzido a quatro fatores que continham um alta proporção de variabilidade presente no conjunto de dados original (81,9%), deixando apenas 18,1% da variação para os fatores específicos. Outros estudos, visando estabelecer inter-relações entre peso e medidas corporais (KESKIN; DASKIRAN; KOR, 2007; ÖNK; SARI; GÜRCAN, 2018; TAHTALI, 2019), também relataram alta proporção de variabilidade presente nas variáveis originais, com valores variando de 73,7% a 78,9%.

2. ADEQUAÇÃO DE DADOS PARA ANÁLISE FATORIAL

A medida de adequação da amostragem de Kaiser-Meyer-Olkin (KMO) e o teste de esfericidade de Bartlett são aplicados para determinar se a matriz de correlação é uma matriz de identidade. O teste de KMO é uma medida que calcula tanto para toda a matriz de correlação quanto para cada variável individual, ou seja, não avalia apenas as correlações, mas os padrões entre as variáveis (HAIR JR. *et al.*, 2009). Esse índice varia de 0 a 1, para proceder com análise fatorial, o banco de dados analisado precisa ter um valor de $KMO > 0,50$, $KMO < 0,50$, que é considerado inaceitável. O teste de Bartlett avalia a significância geral de todas as correlações em uma matriz de correlação. Nesse caso, se a hipótese nula não for rejeitada, as variáveis não são correlacionadas e inadequadas para serem descritas em fatores ou variáveis latentes (MARQUES *et al.*, 2020).

Para estimar os parâmetros dos fatores existem vários métodos; Howard (2016) destaca como principais a análise de componentes principais, a análise de fatores comuns e máxima verossimilhança. O método de verossimilhança assume normalidade multivariada dos dados (MARQUES *et al.*, 2020), ou seja, requer que a matriz de covariância seja positiva definida (não singular). Sua escolha é adotada quando se pretende explicar a estrutura latente da matriz de correlações (FÁVERO *et al.*, 2009). Por outro lado, a análise de componentes principais pode ser aplicada quando a matriz de covariância é singular (MARQUES *et al.*, 2020). Quando a distribuição é singular, os componentes podem ser expressos como combinações lineares de várias variáveis independentes baseada em autovalores e autovetores. O mesmo fato ocorre com

a análise de fatores comuns, porém o objetivo primário é identificar fatores latentes que reflitam o que as variáveis têm em comum (HAIR JR *et al.*, 2009).

Antes de realizar a análise fatorial, também é importante determinar o número adequado de fatores. O objetivo da extração do número de fatores é determinar a quantidade de fatores que melhor representa o padrão de correlação entre as variáveis observadas (FIGUEIREDO FILHO; SILVA JÚNIOR, 2010).

Existem muitas regras e abordagens de extração, incluindo: critérios de Kaiser (autovalor ≥ 1), Teste de *Scree*, percentual de variância e análise paralela (WILLIAMS; ONSMAN; BROWN, 2010). A maioria desses métodos cria pontos de corte de fatores por meio da análise de autovalores, que são valores numéricos que representam as “variações nas variáveis medidas contabilizadas por cada um dos fatores comuns”, sendo os autovalores são calculados somando as cargas dos fatores ao quadrado (HOWARD, 2016).

O critério de Kaiser escolhe o número de fatores a reter, em função do número dos autovalores (autovalor ≥ 1), esses autovalores, mostram a variância explicada por cada fator, ou seja, quanto cada fator consegue explicar a variância total. De acordo Fávero *et al.* (2009), a escolha dos componentes que apresentam autovalores ≥ 1 decorre do fato que trabalhando com variáveis padronizadas, com média igual a zero e variância igual a um, logo, somente os fatores que possuem autovalores ≥ 1 serão significativos. Por outro lado, o Teste *Scree*, assim como a análise paralela, é um procedimento gráfico que plota os autovalores em função dos fatores extraídos, ao qual o número de fatores é escolhido na parte curva da plotagem, considerando os fatores que mais contribuem para a explicação da variabilidade dos dados, descartando-se os fatores da parte linear da plotagem (MARQUES *et al.*, 2020).

Uma vez escolhido o número de fatores, as cargas fatoriais das variáveis precisam ser interpretadas; no entanto, os resultados iniciais são difíceis de interpretação. Em virtude disso, utilizam-se as técnicas de rotações. As técnicas de rotações age maximizando as altas cargas fatoriais e minimizando as baixas cargas fatoriais, produzindo, assim, uma interpretação mais simplificada dos dados (WILLIAMS; ONSMAN; BROWN, 2010). Nesse contexto, existem duas técnicas comuns de rotações: a rotação oblíqua e a rotação ortogonal. Os pesquisadores têm vários métodos para escolher entre essas duas respectivas rotações. Para os métodos rotacionais ortogonais, destacam-se o Varimax, o Quartimax e o Equamax. Para os métodos rotacionais oblíquos destacam-se Oblimin e o Promax. As rotações ortogonais não permitem que os fatores rotacionados resultantes sejam correlacionados; alternativamente, as rotações oblíquas permitem correlacionar os fatores rotacionados resultantes (HOWARD, 2016).

A rotação Varimax é, de longe, a rotação ortogonal mais utilizada, provavelmente porque é o padrão em muitos pacotes de *software*, mas também porque foi desenvolvido como uma melhoria incremental em métodos anteriores, Quartimax e Equamax (OSBORNE, 2015). A rotação Varimax procura aumentar as variações das cargas fatoriais, resultando em cargas fatoriais altas e baixas; isso geralmente é preferível, pois as variáveis claramente pesam ou não pesam em cada fator (HOWARD, 2016). Para o método de rotação oblíqua, Osborne (2015) relatou que não a um amplamente preferido, ambos tendem a produzir resultados semelhantes, e parece geralmente bom para usar as configurações padrões nos pacotes de *software*.

Como base no exposto, pode surgir o seguinte questionamento: rotações ortogonais e oblíquas produzem diferenças visíveis? Para essa pergunta, Osborne (2015) esclareceu que rotações ortogonais e oblíquas produzirão respostas praticamente idênticas no caso improvável de os fatores serem perfeitamente não correlacionados. Para Howard (2016), a escolha da rotação deve ser guiada por uma teoria *a priori*; nesse caso, se o pesquisador acredita que os fatores resultantes não estão correlacionados, deve executar uma rotação ortogonal. Por outro lado, se o pesquisador acredita que os fatores resultantes estão correlacionados, deve executar uma rotação oblíqua.

Tradicionalmente, os pesquisadores são orientados à rotação ortogonal, porque os fatores não correlacionados são mais facilmente interpretáveis e são geralmente a configuração padrão na maioria dos pacotes da computação estatística. Porém, não parece haver uma razão convincente para sua utilização; nas ciências, de um modo geral, esperamos correlações entre fatores, uma vez que o comportamento raramente é particionado em unidades ordinalmente que funcionam independentemente um do outro (OSBORNE, 2015). Independentemente de qual método de rotação é usado, os objetivos são fornecer uma interpretação mais fácil dos resultados e produzir resultados mais parcimoniosa (WILLIAMS; ONSMAN; BROWN, 2010).

Em relação à carga fatorial, ela representa a correlação entre uma variável original e seu fator. Ao interpretar os fatores, a análise fatorial também determina a extensão em que cada variável representa em cada fator, por meio das cargas fatoriais (HOWARD, 2016). Nesse sentido, Hair Jr. (2009) sugere especificar cargas fatoriais consideradas significativas com base nos diferentes tamanhos de amostras (Tabela 1).

Tabela 1. Diretrizes para identificação de cargas fatoriais significativas com base no tamanho de amostra.

Carga fatorial	Tamanho da amostra necessário para significância
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

(HAIR JR. *et al.*, 2009)

Podemos observar que à medida que o número de variáveis em análise diminui, o nível aceitável para considerar uma carga significativa aumenta. Assim, essas orientações devem ser usadas como ponto de partida na interpretação das cargas fatoriais, sendo as cargas maiores consideradas significativas e acrescentadas à interpretação com base em outras considerações (HAIR JR. *et al.*, 2009).

Uma vez que todas as cargas significantes tenham sido identificadas, o próximo passo é procurar por variáveis que não sejam adequadamente explicadas pela análise fatorial. Com isso, é necessário examinar a comunalidade de cada variável, que são as quantidades das variâncias (correlações) de cada variável explicada pelos fatores. Hair Jr. (2009) exemplifica que podemos especificar que pelo menos metade da variância de cada variável deve ser levada em conta, ou seja, todas as variáveis com comunalidades menores que 0,50 como não tendo explicação suficiente.

3. USO DOS ESCORES FATORIAIS NA ANÁLISE DE REGRESSÃO MÚLTIPLA

A análise de regressão múltipla é uma técnica clássica comumente usada para determinar a força da relação entre uma variável dependente contínua e duas ou mais variáveis independentes (EKIZ *et al.*, 2020). No entanto, a análise de regressão múltipla, um dos pressupostos para interpretação com base no método dos mínimos quadrados, é a falta de relação significativa entre as variáveis independentes (ausência de multicolinearidade) (TAHTALI, 2019). Porém, a multicolinearidade pode surgir entre as variáveis independentes que são usadas para prever a variável dependente nos resultados da análise. Se houver correlações altas ($> 0,80$) entre as variáveis independentes, as análises de regressão múltipla

podem produzir estimativas enviesadas devido ao problema de multicolinearidade (ALI *et al.*, 2015).

A análise fatorial, mediante os escores fatoriais visa identificar a estrutura subjacente a uma matriz de dados, pela qual reduz um sistema complexo de correlações em dimensões menores por meio da extração de variáveis latentes (não observáveis). No geral, a análise fatorial é um método que explica uma estrutura de muitas variáveis originais correlacionadas em um menor número de novas variáveis (fatores). Esses fatores são independentes entre si, ou seja, não são correlacionados; porém, dentro de cada fator as cargas fatoriais estão correlacionadas. Daskiran *et al.* (2017) descreveram que a utilização dos escores fatoriais da análise fatorial multivariada pode ser útil para eliminar problemas de multicolinearidade entre um grande número de variáveis. Acrescentou ainda que esse tipo de abordagem facilita o ajuste de vários modelos de regressão, removendo o efeito indireto da explicação relacionada às variáveis.

Com isso Keskim *et al.* (2007) estudaram a relação entre o peso da carcaça e 10 medidas corporais (peso de abate, altura da cernelha, comprimento do corpo, profundidade, largura e circunferência do peito, circunferência, largura e comprimento da perna, e largura da garupa), que foram examinados por meio da análise fatorial e regressão múltipla. Neste estudo foi possível observar a seleção de três fatores como variáveis independentes para o modelo de regressão múltipla; com isso 83,9% da variação do peso da carcaça foram explicados por esses três fatores.

Cadavez (2009) também concluiu que a análise fatorial pode ser um procedimento eficiente para evitar a multicolinearidade entre variáveis independentes e desenvolver modelos mais estáveis. Neste estudo supracitado, medidas de carcaças foram usadas para prever o rendimento de carcaça magra em cordeiros (Churro Bragançana e Suffolk), possibilitando a seleção de quatro variáveis latentes que foram responsáveis por 95,4% da variância comum. Khan *et al.* (2014) avaliaram ovinos machos e fêmeas da raça Harnai para estimar o peso corporal mediante várias medidas corporais. O uso dos escores fatoriais de quatro variáveis latentes na análise de regressão possibilitou em coeficientes de determinação de 87,8 e 92,0% para machos e fêmeas, respectivamente.

Daskiran *et al.* (2017) relataram um caso típico de multicolinearidade em um estudo com cabras leiteiras da raça Norduz para determinar a relação entre a produção diária de leite e as características do úbere. O uso da análise de regressão das variáveis originais, embora o coeficiente de determinação do modelo encontrado tenha sido de 74,8%, e estatisticamente significativo, mostrou que nenhum dos coeficientes das variáveis foram significativos. Por

outro lado, quando os escores fatoriais foram usados na análise de regressão, o estudo descobriu que dois fatores tiveram efeito estatístico significativo na produção diária de leite e esses fatores juntos responderam por 78,6% da variação da produção diária de leite. Por outro lado, Çelik *et al.* (2018) avaliaram a influência dos pesos das partes da carcaça (coxa, peito, asa, dorso, moela e pés) sobre o peso total de carcaças em perus. Nesse estudo foram relatados problemas de multicolinearidade para as variáveis peso da coxa e asa, por apresentarem valores de Fator de inflação da variância (VIF) maiores que 10 (15,64 e 12,23, respectivamente). No entanto, quando os escores fatoriais foram obtidos a partir da análise fatorial foram usadas como variáveis independentes para estimar o peso da carcaça em perus, o problema da multicolinearidade foi resolvido e foi encontrado $VIF = 1$. Os dois escores fatoriais usados no modelo explicaram 96,6% da variação total do peso da carcaça, sendo considerados adequados para prever o peso de carcaça em perus.

Tahtali (2019) realizou um estudo com base em medidas corporais para prever o peso corporal em cordeiros da raça Romanov. Nesse estudo também foi observada uma multicolinearidade entre o peso da garupa e a altura de cernelha ($VIF > 10$). Para resolver o problema de multicolinearidade e para uma análise de regressão confiável, os escores fatoriais obtidos da análise fatorial foram usados como variáveis independentes na análise de regressão múltipla. Com isso, os coeficientes dos escores fatoriais dos dois primeiros fatores obtidos com a análise fatorial multivariada foram usados com variáveis independentes, de acordo os resultados os efeitos dos dois fatores foram estatisticamente significativos ($P < 0,001$) e a multicolinearidade entre as variáveis independentes originais foi eliminada usando os escores fatoriais ($VIF = 1,0$). Além disso, os escores dos fatores utilizados no modelo explicaram 75,4% da variância total dos pesos corporais dos cordeiros.

O VIF é um índice que indica o quanto a variância de um coeficiente de regressão é aumentado devido à colinearidade. Nesse sentido, não existe um valor de corte ou método formal para determinar quando um VIF é muito grande, sugestões típicas para um ponto de corte são 5 ou 10 (CRANEY; SURLLES, 2002). No entanto, existe um senso comum de que valores de VIF superiores a 5 sugerem que os coeficientes de regressão são mal estimados devido à multicolinearidade severa. Na Tabela 2 são apresentadas as diretrizes sugeridas por Kazemi *et al.* (2013) para interpretação os valores de VIF.

Tabela 2. Diretrizes para interpretar os valores de VIF (Fator de inflação da variância).

VIF	Status das preditoras
VIF = 1	Não correlacionado
$1 < \text{VIF} < 5$	Moderadamente correlacionado
VIF > 5	Altamente correlacionado

(KAZEMI *et al.*, 2013)

4. ANÁLISE DISCRIMINANTE MULTIVARIADA

A análise discriminante multivariada é uma técnica estatística que leva em conta as diferentes variáveis de um objeto e funciona encontrando as chamadas funções discriminantes importantes, de forma que as diferenças entre os grupos predefinidos sejam maximizado (DU; SUN, 2008). Com descrito por Chandraratne *et al.* (2006), a análise discriminante encontra um conjunto de combinações lineares das variáveis, cujos valores são o mais próximo possível dentro dos grupos e mais distante possível entre grupos. Um conhecimento prévio dos grupos predefinidos das amostras testadas é um pré-requisito primordial para a aplicação da análise discriminante multivariada, para diferenciação entre os grupos (ELMASRY *et al.*, 2011).

A análise discriminante multivariada também é aplicada para prever a qual grupo uma observação pertence, com base no conhecimento das variáveis quantitativas em um conjunto de combinações lineares, sendo que as funções discriminantes lineares ou quadráticas podem ser usadas para dados com distribuições normais (PARK *et al.*, 2002). Na análise discriminante linear (LDA) supõe-se que os grupos tenham matrizes de covariância iguais e diferem apenas em suas médias, enquanto a análise discriminante quadrática (QDA), as matrizes de covariância dos grupos não são consideradas iguais (MAHARAJ; ALONSO, 2014).

Uma análise das características de textura, com base em matrizes de co-ocorrência, foi realizada para discriminar as carcaças de aves prejudiciais e saudáveis (PARK *et al.*, 2002). Nesse estudo, modelos discriminantes foram desenvolvidos para separar carcaças prejudiciais de carcaças saudáveis. A precisão média do modelo LDA para identificar carcaças prejudiciais foi de 90,5%, porém, a precisão para a identificação de carcaças saudáveis foi baixa (menos de 51,5%). Nesse contexto, os modelos lineares têm maior dificuldade em prever carcaças de aves saudáveis. Quando o modelo QDA foi aplicado, a precisão foi de 75% e 62% para identificar carcaças saudáveis e prejudiciais, respectivamente. Com isso, os autores concluíram que os modelos quadráticos foram menos precisos que os modelos lineares e o desenvolvimento de um modelo discriminante adicional utilizando outros parâmetros na análise de características texturais deve ser conduzido para melhorar a classificação de carcaças de aves.

Visando prever o grau de maciez em carcaça de cordeiros utilizando recursos de imagens extraídos do músculo *longissimus dorsi*, Chandrarante *et al.* (2006b) utilizaram a análise discriminante multivariada como método de classificação. A análise discriminante multivariada foi aplicada usando funções discriminantes lineares e quadráticas e, em todos os casos, as funções discriminantes lineares produziram melhor classificação do que as funções discriminantes quadráticas. De acordo os autores, a classificação mostrou resultados encorajadores, indicando que os dados extraídos de imagens do músculo *longissimus dorsi* de cordeiros pode ser efetivamente utilizado para prever o grau de maciez de carcaças de cordeiros, uma vez que apresentou precisão de 85%. Sendo importante destacar que dos 16 grupos de produtos avaliados a maioria foram classificados com precisão de 99% a 100% em seu grupo de origem.

Kamruzzaman *et al.* (2011) também utilizaram a LDA para classificar os músculos de cordeiros por intermédio de imagens HSI (Hue Saturation Intensity). O modelo de LDA desenvolvido para reconhecimento muscular foi capaz de identificar todas as amostras dos músculos (*Semitendinosus*, *Longissimus dorsi* e *Psoas Major*) com precisão de classificação de 100%. Por outro lado, a análise discriminante multivariada mostrou-se também eficiente para desenvolver um método não destrutivo para classificar a maciez da carne cozida (14 dias *pós-mortem*) em estudo realizado por Cluff *et al.* (2013) com análise do músculo *longissimus dorsi* em bovinos. O modelo LDA desenvolvido foi capaz de classificar amostras resistentes e tenras com 83,3% e 75,0% de precisão, respectivamente. Em um estudo semelhante, Naganathan *et al.* (2015) desenvolveu um modelo para prever a maciez da carcaça bovina, ao qual imagens HSI coletaram simultaneamente a estrutura muscular e informações bioquímicas que têm altos graus de relação com a maciez da carne. O modelo LDA desenvolvido mediante o músculo *Longissimus dorsi* da carne bovina adquirida em dois dias *pós-mortem* poderia prever a maciez da carne cozida em 14 dias, com precisão de 92,9%.

A análise discriminante multivariada também tem sido destinada para avaliar alimentos industrializados. Nesse sentido, um modelo foi desenvolvido para avaliar a qualidade de uma variedade de produtos resfriados oriundos da carne suína e de aves. A análise discriminante foi aplicada com objetivo de encontrar um conjunto de variáveis que minimizassem o erro durante a discriminação dos produtos de suínos e aves pertencentes à mesma categoria de processamento (ZAPOTOCZNY; SZCZYPIŃSKI; DASZKIEWICZ, 2016). E com base nas variáveis mais altamente discriminantes, os produtos oriundos das aves foram discriminados com precisão de 78% a 100%, ou seja, variou de acordo o tipo de produtos onde geralmente

uns são mais difíceis de classificar comparado a outros. Para os produtos advindos da carne suína, os produtos avaliados foram classificados com precisão de 82% a 100%.

4.1 Vizinho mais próximo (K-NN)

Como relatado anteriormente, para a aplicação da análise discriminante multivariada linear e quadrática, os dados devem atender o pressuposto de normalidade. Quando esse pressuposto não é atendido, pesquisadores têm buscado procedimentos alternativos, principalmente a utilização de métodos não paramétricos, que não precisam do atendimento de normalidade, dentre eles o vizinho mais próximo (K-NN). Como descrito por Du e Sun (2008), o K-NN é uma técnica de classificação não paramétrica realizada atribuindo um caso desconhecido à classe mais frequentemente representada entre as amostras mais próximas. Esse método assume que todas as instâncias correspondem a pontos no espaço n -dimensional (BARBON *et al.*, 2016) e, para classificar uma nova instância, o classificador K-NN seleciona k vizinhos mais próximos do conjunto de dados de acordo com uma distância métrica (TAHERI-GARAVAND *et al.*, 2019). Adeniyi, Wei e Yongquan (2016), por exemplo, cita a distância euclidiana para medir as semelhanças entre os dados em treinamento. Depois que k vizinhos são encontrados, um valor médio é calculado entre vizinhos e atribuídos com valor de previsão a uma instância desconhecida (BARBON *et al.*, 2017).

Por ser classificado com base na classe de seus vizinhos mais próximos, na maioria das vezes, mais de um vizinho é levado em consideração, portanto, o nome K-NN o “K” indica o número de vizinhos usados para determinar a taxa de classificação (ADENIYI; WEI; YONGQUAN, 2016). Taheri-Garavand *et al.* (2019), em seu estudo, experimentaram diferentes valores de K (1 a 10) para determinar a frescura do peixe e concluíram que 8-NN obteve o melhor desempenho com precisão de classificação de 90,48%.

O classificador K-NN foi utilizado para a categorização dos vários estágios de cozimento da carne moída usada na fabricação de hambúrgueres (O’FARRELL *et al.*, 2005). Os diferentes estágios de cozimento foram definidos como: cru, claro, ideal e escuro. E para a distinção dos diferentes estágios 1-KK foi o k que obteve o melhor resultado. A precisão de classificação foi de 100% para os estágios cru, claro e ótimo. O estágio escuro obteve apenas 60% de precisão, sendo que 40% foram classificados como ótimo. Como o cozimento é um processo contínuo, os estágios de cozimento não são mutuamente exclusivos e existe uma forte correlação entre eles, isso significa que a classificação dos espectros limítrofes pode ser difíceis, assim como limites de cada estágio de cozimento não serem precisos (O’FARRELL *et al.*, 2005).

Usando o sistema de visão computacional para classificação de marmoreio da carne bovina e suína, Barbon *et al.* (2017) utilizaram o algoritmo K-NN como método de classificação. E nessa abordagem a distância euclidiana foi usada para atribuir valores de previsão a uma instância desconhecida e diferentes valores de k foram testados para determinar a taxa de classificação. Nesse estudo, mostraram que o melhor valor de k resultou em 3-NN tanto para o conjunto de dados da carne bovina e suína. Esse resultado pode ser observado de forma visual (Figura 1), que é destacado pela linha vertical, que mostra a precisão média calculada sobre os 20 valores de k testados.

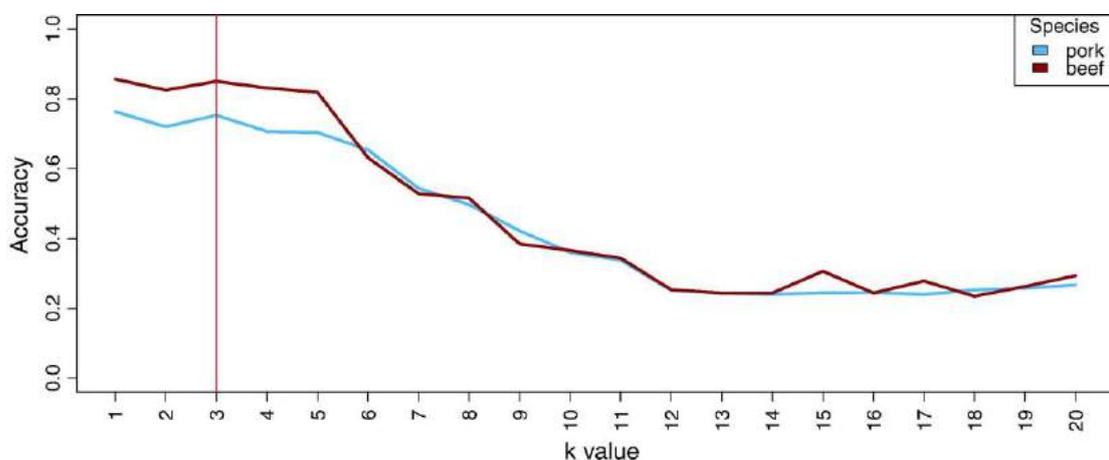


Figura 1. Comparação de vizinhos próximos (k) avaliados em experimentos com carne suína e bovina (BARBON *et al.*, 2017).

Esses resultados resultaram em valores de precisão para classificação de 81,59 e 76,14% para carne bovina e suína, respectivamente. Os resultados obtidos demonstraram que a abordagem K-NN pode identificar corretamente a pontuação de marmoreio tanto da carne bovina quanto da suína (BARBON *et al.*, 2017).

5. REFERÊNCIAS

- ADENIYI, D. A.; WEI, Z.; YONGQUAN, Y. Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. **Applied Computing and Informatics**, v. 12, n. 1, p. 90–108, 2016.
- ALI, M. *et al.* Comparison of artificial neural network and decision tree algorithms used for predicting live weight at post weaning period from some biometrical characteristics in harnai sheep. **Pakistan Journal of Zoology**, v. 47, n. 6, p. 1579–1585, 2015.
- ALKARKHI, A. F.; ALQARAGHULI, W. A. A. Discriminant Analysis. In: **Applied Statistics for Environmental Science with R**, p. 173–190, 2020.
- ALVES, M. F.; LOTUFO, A. D. P.; LOPES, M. L. M. Seleção de variáveis stepwise aplicadas

- em redes neurais artificiais para previsão de demanda de cargas elétricas. **Proceeding Series of the Brazilian Society of Applied and Computational Mathematics**, v. 1, n. 1, p. 1–6, 2013.
- ARSALANE, A. *et al.* Beef and horse meat discrimination and storage time classification using a portable device based on DSP and PCA method. **International Journal of Intelligent Enterprise**, v. 4, n. 1/2, p. 58, 2017.
- ARSALANE, A. *et al.* An embedded system based on DSP platform and PCA-SVM algorithms for rapid beef meat freshness prediction and identification. **Computers and Electronics in Agriculture**, v. 152, p. 385–392, 2018.
- BARBON, A. P. A. C. *et al.* Storage time prediction of pork by Computational Intelligence. **Computers and Electronics in Agriculture**, v. 127, p. 368–375, 2016.
- BARBON, A. P. A. DA C. *et al.* Development of a flexible Computer Vision System for marbling classification. **Computers and Electronics in Agriculture**, v. 142, p. 536–544, 2017.
- BUETTNER, F. *et al.* f-scLVM: Scalable and versatile factor analysis for single-cell RNA-seq. **Genome Biology**, v. 18, n. 1, p. 1–13, 2017.
- CADAVEZ, V. Interrelationships between carcass weight, dimensions and tissues thickness measurements on Churro Galego Bragançano and Suffolk Lambs. **Revista Portuguesa de Zootecnia**, v. 1, n. 1, p. 1–9, 2021.
- _____. Prediction of lean meat proportion of lamb carcasses. **Archiva Zootechnica**, v. 12, n. 4, p. 46–58, 2009.
- ÇELİK, Ş. *et al.* Analysis of variables affecting carcass weight of white Turkeys by regression analysis based on factor analysis scores and ridge regression. **Revista Brasileira de Ciência Avícola**, v. 20, n. 2, p. 273–280, 2018.
- CHANDRARATNE, M. R. *et al.* Prediction of lamb tenderness using image surface texture features. **Journal of Food Engineering**, v. 77, n. 3, p. 492–499, 2006.
- CLUFF, K. *et al.* Optical scattering with hyperspectral imaging to classify longissimus dorsi muscle based on beef tenderness using multivariate modeling. **Meat Science**, v. 95, n. 1, p. 42–50, 2013.
- CONTE, G. *et al.* Investigating mutual relationship among milk fatty acids by multivariate factor analysis in dairy cows. **Livestock Science**, v. 188, p. 124–132, 2016.
- CORREDDU, F. *et al.* Use of multivariate factor analysis to characterize the fatty acid profile of buffalo milk. **Journal of Food Composition and Analysis**, v. 60, p. 25–31, 2017.
- CRANEY, T. A.; SURLES, J. G. Model-dependent variance inflation factor cutoff values. **Quality Engineering**, v. 14, n. 3, p. 391–403, 2002.
- DASKIRAN, I.; KESKIN, S.; BINGOL, M. Usability of the factor analysis scores in multiple linear regression analyses for the prediction of daily milk yield in Norduz Goats. **Journal of Agricultural Science and Technology**, v. 19, p. 1507–1515, 2017.
- DU, C.; SUN, D. Object Classification Methods. In: **Computer Vision Technology for Food Quality Evaluation**, p. 81–107, 2008.
- EKIZ, B. *et al.* Comparison of the decision tree, artificial neural network and multiple regression methods for prediction of carcass tissues composition of goat kids. **Meat Science**, v. 161, p. 108011, 2020.
- ELMASRY, G. *et al.* Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. **Journal of Food Engineering**, v. 103, n. 3, p. 333–344, 2011.

- EYDURAN, E. *et al.* Carcass weight estimation from some morphological traits of Capoeira capoeira (Güldenstädt, 1772) using factor scores in multiple regression analysis. **Pakistan Journal of Statistics**, v. 28, n. 1, p. 159–165, 2012.
- FÁVERO, L. P. *et al.* Análise Fatorial. In: **Análise de dados: modelagem multivariada para tomada de decisões**, p. 235–265, 2009.
- FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Visão além do alcance: Uma introdução à análise fatorial. **Opinião Pública**, v. 16, n. 1, p. 160–185, 2010.
- GOMES, H. F. B. *et al.* Common factors method to predict the carcass composition tissue in kid goats. **Revista Brasileira de Zootecnia**, v. 42, n. 3, p. 193–203, 2013.
- HAIR JR, J. *et al.* **Análise Multivariada de Dados**. 6 ed. Porto Alegre: Bookman, 2009.
- HOQUE, A. S. M. . *et al.* Exploratory factor analysis of entrepreneurial orientation in the context of Bangladeshi Small and Medium Enterprise (SMES). **European Journal of Management and Marketing Studies**, v. 3, n. 2, p. 81–94, 2018.
- HOSSEINPOUR, S.; ILKHCHI, A. H.; AGHBASHLO, M. An intelligent machine vision-based smartphone app for beef quality evaluation. **Journal of Food Engineering**, v. 248, p. 9–22, 2019.
- HOWARD, M. C. A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? **International Journal of Human-Computer Interaction**, v. 32, n. 1, p. 51–62, 2016.
- JEON, H. *et al.* Discrimination of origin of sesame oils using fatty acid and lignan profiles in combination with canonical discriminant analysis. **JAACS, Journal of the American Oil Chemists' Society**, v. 90, n. 3, p. 337–347, 2013.
- KAMRUZZAMAN, M. *et al.* Application of NIR hyperspectral imaging for discrimination of lamb muscles. **Journal of Food Engineering**, v. 104, n. 3, p. 332–340, 2011.
- KAMRUZZAMAN, M. *et al.* Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. **Innovative Food Science and Emerging Technologies**, v. 16, p. 316–325, 2012.
- KAMRUZZAMAN, M. *et al.* Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. **Talanta**, v. 103, p. 130–136, 2013.
- KAZEMI, A. *et al.* Optimal power quality monitor placement using genetic algorithm and Mallow's Cp. **International Journal of Electrical Power and Energy Systems**, v. 53, n. 1, p. 564–575, 2013.
- KESKIN, S.; DASKIRAN, I.; KOR, A. Factor analysis scores in a multiple linear regression model for the prediction of carcass weight in akkeci kids. **Journal of Applied Animal Research**, v. 31, n. 2, p. 201–204, 2007.
- KHAN, M. A. *et al.* Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. **Journal of Animal and Plant Sciences**, v. 24, n. 1, p. 120–126, 2014.
- LAWLEY, D. N.; MAXWELL, A. E. Factor Analysis as a Statistical Method. **Journal of the Royal Statistical Society. Series D (The Statistician)**, v. 12, n. 3, p. 209–229, 1962.
- MACCIOTTA, N. P. P. *et al.* A multivariate approach to modeling shapes of individual lactation curves in cattle. **Journal of Dairy Science**, v. 87, n. 4, p. 1092–1098, 2004.

- MACCIOTTA, N. P. P. *et al.* Use of multivariate factor analysis to define new indicator variables for milk composition and coagulation properties in Brown Swiss cows. **Journal of Dairy Science**, v. 95, n. 12, p. 7346–7354, 2012.
- MAHARAJ, E. A.; ALONSO, A. M. Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals. **Computational Statistics and Data Analysis**, v. 70, p. 67–87, 2014.
- MANCA, M. G. *et al.* Derivation of multivariate indices of milk composition, coagulation properties, and individual cheese yield in dairy sheep. **Journal of Dairy Science**, v. 99, n. 6, p. 4547–4557, 2016.
- MARQUES, R. A. M. *et al.* Multivariate GR&R through factor analysis. **Measurement: Journal of the International Measurement Confederation**, v. 151, p. 3–17, 2020.
- MELE, M. *et al.* Multivariate factor analysis of detailed milk fatty acid profile: Effects of dairy system, feeding, herd, parity, and stage of lactation. **Journal of Dairy Science**, v. 99, n. 12, p. 9820–9833, 2016.
- NAGANATHAN, G. K. *et al.* A prototype on-line AOTF hyperspectral image acquisition system for tenderness assessment of beef carcasses. **Journal of Food Engineering**, v. 154, p. 1–9, 2015.
- O’FARRELL, M. *et al.* Comparison of k-NN and neural network methods in the classification of spectral data from an optical fibre-based sensor system used for quality control in the food industry. **Sensors and Actuators, B: Chemical**, v. 111–112, p. 354–362, 2005.
- ÖNK, K.; SARI, M.; GÜRCAN, İ. S. Estimation of live weights at the beginning and the end of grazing season in Tuj lambs via scores of factor analysis. **Ankara Universitesi Veteriner Fakültesi Dergisi**, v. 65, n. 3, p. 261–266, 2018.
- OSBORNE, J. W. What is rotating in exploratory factor analysis? **Practical Assessment, Research and Evaluation**, v. 20, n. 2, p. 1–7, 2015.
- PAPADOPOULOU, O. S. *et al.* Sensory and microbiological quality assessment of beef fillets using a portable electronic nose in tandem with support vector machine analysis. **Food Research International**, v. 50, n. 1, p. 241–249, 2013.
- PARK, B. *et al.* Discriminant analysis of dual-wavelength spectral images for classifying poultry carcasses. **Computers and Electronics in Agriculture**, v. 33, n. 3, p. 219–231, 2002.
- PATRIARCA, R. *et al.* Resilience engineering: Current status of the research and future challenges. **Safety Science**, v. 102, p. 79–100, 2018.
- TAHERI-GARAVAND, A. *et al.* Real-time nondestructive monitoring of Common Carp Fish freshness using robust vision-based intelligent modeling approaches. **Computers and Electronics in Agriculture**, v. 159, p. 16–27, 2019.
- TAHTALI, Y. Use of factor scores in multiple regression analysis for estimation of body weight by certain body measurements in Romanov Lambs. **PeerJ**, v. 2019, n. 8, p. 1–11, 2019.
- TODARO, M.; SCATASSA, M. L.; GIACCONE, P. Multivariate factor analysis of Girgentana goat milk composition. **Italian Journal of Animal Science**, v. 4, n. 4, p. 403–410, 2005.
- TURINI, L. *et al.* Multivariate factor analysis of milk fatty acid composition in relation to the somatic cell count of single udder quarters. **Journal of Dairy Science**, v. 103, n. 8, p. 7392–7406, 2020.
- VICTOR, V. *et al.* Factors influencing consumer behavior and prospective purchase decisions

in a dynamic pricing environment-an exploratory factor analysis approach. **Social Sciences**, v. 7, n. 9, p. 2–14, 2018.

WILLIAMS, M. B.; ONSMAN, A.; BROWN, T. Exploratory factor analysis: A five-step guide for novices. **Journal of Emergency Primary Health Care**, v. 8, n. 3, p. 1–13, 2010.

ZAPOTOCZNY, P.; SZCZYPIŃSKI, P. M.; DASZKIEWICZ, T. Evaluation of the quality of cold meats by computer-assisted image analysis. **LWT - Food Science and Technology**, v. 67, p. 37–49, 2016.

CAPÍTULO II

**Adequação da análise fatorial e dos métodos de rotação para a
caracterização de carcaças de ovinos da raça Santa Inês**

ADEQUAÇÃO DA ANÁLISE FATORIAL E DOS MÉTODOS DE ROTAÇÃO PARA A CARACTERIZAÇÃO DE CARCAÇAS DE OVINOS DA RAÇA SANTA INÊS

RESUMO – A análise fatorial é uma técnica multivariada que tem como finalidade reduzir o número de variáveis que explicam a variação total dos dados em fatores e examinar a estrutura ou relações entre as variáveis. No entanto, a execução da análise fatorial requer tomada de decisões tanto estatísticas quanto metodológicas, para validação da técnica e facilitar a interpretação dos resultados. Em virtude disso, objetivou-se avaliar a adequação da análise fatorial e dos métodos rotacionais para caracterização de carcaças de ovinos da raça Santa Inês. O banco de dados foi composto por informações de características de carcaças de 122 ovinos não castrados da raça Santa Inês, submetidos ao regime de confinamento. O teste de esfericidade de Bartlett e KMO (Kaiser-Meyer-Olkin) indicou que as variáveis de carcaças avaliadas foram adequadas para aplicação da análise fatorial. Dentre as rotações testadas, a rotação ortogonal Varimax foi a que apresentou estrutura mais simples para interpretação dos fatores construídos.

Palavras-chave: Cortes cárneos. Medidas morfométricas. Promax. Varimax.

ADEQUACY OF FACTOR ANALYSIS AND ROTATION METHODS FOR THE CHARACTERIZATION OF THE SHEEP CARCASSES

ABSTRACT – Factor analysis is a multivariate technique to reduce the number of original variables in factors that explain the total variation of the data, allowing to examine the structure or relations between the variables. However, the execution of factor analysis requires decision making, in statistical and methodological terms, to validate the technique and facilitate the interpretation of the results. Thus, the adequacy of factor analysis and rotation methods was evaluated using, for example, morphometric information of sheep carcass. The database was composed of 122 non-castrated Santa Inês sheep, submitted to a confinement regime. The sphericity test of Bartlett's and KMO (Kaiser-Meyer-Olkin) indicated that the carcass variables evaluated were adequate for the application of factor analysis. Among the tested rotations, the orthogonal Varimax rotation presented the simplest structure for the interpretation of the constructed factors.

Keywords: Biometry. Multivariate analysis. Rotated factors. Varimax.

1. INTRODUÇÃO

A análise fatorial é um procedimento estatístico multivariado que atua em diferentes escalas. Como descrito por Williams, Onsman e Brown (2010), em primeiro lugar reduz muitas variáveis em um conjunto menor de variáveis (também conhecidas como fatores), e em segundo lugar, estabelece dimensões subjacentes entre variáveis medidas e construções latentes, ou seja, aquelas relações não observáveis. Entretanto, ao executar a técnica de análise fatorial, várias decisões estatísticas e metodológicas são necessárias. Dentre essas podemos destacar as técnicas de inspeção de dados, método de extração dos fatores, método de retenção dos fatores, método de rotação dos fatores e nível de significância das cargas fatoriais. Infelizmente, as

escolhas para tais decisões geralmente não são claras, fazendo com que pesquisadores obtenham resultados imprecisos (HOWARD, 2016).

A exemplo disso, podemos destacar os métodos de rotações dos fatores que, tradicionalmente, os pesquisadores são orientados a utilizar a rotação ortogonal. De acordo com Osborne (2015), a justificativa para tal orientação leva em conta que os fatores ortogonais são mais facilmente interpretáveis, e essa medida se justificava em grande parte no século XX, quando a análise fatorial era realizada por cálculos manuais ou poder de computação mais limitado.

Porém, atualmente, esses procedimentos são disponíveis em vários softwares estatísticos, de modo que parece não haver razão plausível para não se testar os demais tipos de rotações e decidir por aquela mais adequada em cada caso. Nas ciências de um modo geral, esperam-se correlações entre os fatores, uma vez que o comportamento entre variáveis raramente é particionado em unidades que funcionam independentemente um do outro (OSBORNE, 2015). Nesse caso, quando os fatores são correlacionados devemos optar pelas rotações oblíquas que considera a existência de correlações entre os fatores.

Para os métodos rotacionais ortogonais, destacam-se o Varimax, o Quartimax e o Equamax. Para os métodos rotacionais oblíquos destacam-se Oblimin direto e o Promax. As rotações ortogonais não permitem que os fatores rotacionados resultantes sejam correlacionados; alternativamente, as rotações oblíquas permitem correlacionar os fatores rotacionados resultantes (HOWARD, 2016).

Na avaliação de carcaça de ovinos são considerados um número elevado de variáveis é possível que muitas delas contribuam pouco para a discriminação dos indivíduos avaliados, por serem invariantes entre estes indivíduos, redundantes em virtude das correlações, ou ainda, pelo fato de uma variável ser uma combinação linear de outras (BARBOSA *et al.*, 2006), dificultando a interpretação dos resultados.

Dado o exposto, objetivou-se avaliar a adequação da análise fatorial e dos métodos rotacionais para caracterização de carcaças de ovinos da raça Santa Inês.

2. MATERIAL E MÉTODOS

O banco de dados continha informações de 122 ovinos não castrados da raça Santa Inês, submetidos ao regime de confinamento. Esse banco de dados é oriundo de três experimentos conduzidos no setor de Caprinocultura e Ovinocultura pertencente ao Centro de Ciências Humanas, Sociais e Agrárias da Universidade Federal da Paraíba, localizado no município de Bananeiras, estado da Paraíba, microrregião do Brejo Paraibano. Do banco de dados total,

foram selecionadas 24 variáveis relacionadas à carcaça dos animais, as quais estão apresentadas na Tabela 1, com respectivas média (μ) e desvio padrão (DP).

Tabela 1. Média (μ) e desvio padrão (DP) das variáveis de carcaças ovinas da raça Santa Inês.

Variáveis	Abreviações	$\mu \pm DP$
Peso de corpo vazio (kg)	PCV	25,29 \pm 3,17
Peso vivo ao abate (kg)	PVA	31,65 \pm 3,67
Peso de carcaça quente (kg)	PCQ	14,91 \pm 1,73
Peso de carcaça fria (kg)	PCF	14,65 \pm 1,67
Rendimento biológico (%)	RB	59,15 \pm 3,57
Rendimento de carcaça quente (%)	RCQ	47,22 \pm 3,09
Rendimento de carcaça fria (%)	RCF	46,41 \pm 3,17
Comprimento externo da carcaça (cm)	CEC	56,19 \pm 2,91
Comprimento interno da carcaça (cm)	CIC	61,18 \pm 4,16
Largura da garupa (cm)	LG	17,62 \pm 3,05
Largura do tórax (cm)	LT	17,51 \pm 3,49
Perímetro da garupa (cm)	PG	55,25 \pm 5,50
Perímetro do pernil (cm)	PP	37,33 \pm 2,74
Perímetro do tórax (cm)	PT	68,23 \pm 3,40
Pernil (kg)	PER	2,24 \pm 0,28
Lombo (kg)	LOM	0,96 \pm 0,17
Costela (kg)	COS	2,14 \pm 0,30
Pescoço (kg)	PES	0,89 \pm 0,30
Paleta (kg)	PAL	1,36 \pm 0,16
Rendimento do pernil (%)	RPER	30,66 \pm 3,23
Rendimento do lombo (%)	RLOM	13,15 \pm 1,88
Rendimento da costela (%)	RCOS	29,30 \pm 3,41
Rendimento do pescoço (%)	RPES	7,35 \pm 1,27
Rendimento da paleta (%)	RPAL	18,68 \pm 2,05

O tamanho da amostra foi analisado de acordo a metodologia de Hair Jr. *et al.* (2009) e Mingoti (2005). Hair Jr. *et al.* (2009) consideram um tamanho amostral adequado $n > 50$ e ter no mínimo cinco vezes mais observações do que o número de variáveis analisadas. Mingoti (2005) leva em consideração a subtração do tamanho da amostra em relação ao número de variáveis ($n - p$), em que n representa o tamanho da amostra e p o número de variáveis respostas analisadas.

A padronização das variáveis foi realizada pela conversão de cada variável em um escore padrão pela subtração da média e desvio-padrão para cada variável. Segue a fórmula abaixo:

$$Z = (X - \mu) / \sigma$$

Em que:

X representa os valores das variáveis;

μ representa o valor médio das variáveis;
 σ representa o desvio-padrão das variáveis.

A necessidade de padronização das variáveis deve-se ao fato de a maioria das medidas de distância ser bastante sensível a diferentes escalas ou magnitudes entre as variáveis (HAIR JR. *et al.*, 2009). A padronização converte variáveis a uma única escala e variabilidade comum.

A matriz de correlação de Pearson $P_{24 \times 24}$ das variáveis originais X_i (PCV, PVA, PCQ, PCF, RB, RCQ, RCF, CEC, CIC, LG, LT, PG, PP, PT, PER, LOM, COS, PES, PAL, RPER, RLOM, RCOS, RPES, RPAL) foi utilizada, com $i = 1, 2, \dots, 24$. A multicolinearidade foi analisada pelos testes estatísticos de adequação da amostragem, o teste de esfericidade de Bartlett e o teste de Kaiser-Meyer-Olkin (KMO). O teste de Bartlett é um teste que avalia a significância geral de todas as correlações em uma matriz de correlação. Se o valor encontrado é estatisticamente significativo ($< 0,05$), indica que existem correlações suficientes para continuar a análise. O teste de KMO é uma medida calculada tanto para toda matriz de correlação quanto para cada variável individual, ou seja, não avalia apenas as correlações, mas os padrões entre as variáveis. Como podemos observar na Tabela 2, esse índice varia de 0 a 1, sendo que, para proceder a análise fatorial, os dados avaliados precisam ter um valor de KMO $> 0,50$.

Tabela 2. Critério de Kaiser-Meyer-Olkin (KMO)

Valor	Grau de adequação da amostra
1,00 – 0,90	Muito boa
0,80 – 0,90	Boa
0,70 – 0,80	Média
0,60 – 0,70	Razoável
0,50 – 0,60	Má
$< 0,50$	Inaceitável

(HAIR JR. *et al.*, 2006)

A normalidade multivariada foi feita pelo teste de Mardia, que se baseia nos desvios de assimetria e curtose. Porém, em vez de fazer a comparação da distribuição de uma variável com uma distribuição normal univariada, o teste multivariado está comparando a distribuição conjunta de várias variáveis com uma distribuição normal multivariada.

A análise de componentes principais foi utilizada como método para retenção dos fatores. O objetivo da análise de componentes principais foi resumir a maior parte da

informação original (variância) em um número mínimo de fatores necessários para explicar o máximo da variância representada pelas variáveis originais.

A determinação dos números de fatores a serem extraídos foi baseada no critério de Kaiser, que considera autovalores ≥ 1 . Estes autovalores, também chamados de eigenvalues, mostram a variância explicada por cada fator, ou seja, quanto cada fator consegue explicar da variância total. De acordo Fávero *et al.* (2009), a escolha dos componentes que apresentam eigenvalues ≥ 1 decorre do fato que trabalhando com variáveis padronizadas, com média igual a zero e variância igual a um; logo, somente os fatores que possuem eigenvalues ≥ 1 serão significativos.

Para facilitar a interpretação das cargas fatoriais é necessária a utilização das técnicas de rotações. Nesse estudo foram utilizadas diferentes técnicas de rotações, com a finalidade de observar qual técnica apresentava melhores resultados para os dados analisados. Foram testadas as rotações ortogonais (Varimax, Quartimax e Equamax) e oblíquas (Oblimin e Promax).

Para análise das cargas fatoriais foram consideradas as diretrizes para identificação das cargas fatoriais significativas com base no tamanho da amostra proposta por Hair Jr. *et al.* (2009), apresentada na Tabela 3.

Tabela 3. Diretrizes para identificação de cargas fatoriais significativas com base em tamanho de amostra.

Carga fatorial	Tamanho da amostra necessário para significância
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

(HAIR JR. *et al.* 2009)

Os procedimentos estatísticos foram realizados pelo programa estatístico SAS[®] OnDemand for Academics. O procedimento PROC CORR foi utilizado para verificação das correlações entre as variáveis. Para realizar o teste de normalidade multivariada Mardia foi utilizado o procedimento %macro multnorm. E o procedimento PROC FACTOR foi usado para

a retenção (análise de componentes principais), extração (critério de Kaiser) e rotações dos fatores (Varimax, Quartimax, Equamax, Oblimin e Promax).

3. RESULTADO E DISCUSSÃO

3.1. Adequação da Análise Fatorial

3.1.1 Tamanho da Amostra

O tamanho da amostra pode interferir substancialmente nos resultados da análise. Claramente, amostras maiores superam amostras menores devido à redução na probabilidade de erros, estimativas mais precisas e melhor generalização dos resultados, sendo que amostra muito pequena, podem gerar erros de inferência (TREIBLMAIER; FILZMOSE, 2010). Várias são as recomendações referente ao tamanho da amostra encontradas na literatura.

Considerando as suposições criadas por Hair Jr. *et al.* (2009) e Mingoti (2005), o banco de dados em estudo demonstrou adequação em relação a tamanho da amostra. O banco de dados foi composto por 120 observações, cujo tamanho da amostra estava de acordo com proposto por Hair Jr. *et al.* (2009), de $n > 50$ e, ter no mínimo cinco vezes mais observações do que o número de variáveis analisadas ($5 * p = 5 * 24$ (número de variáveis) = 120 observações)). Mingoti (2005) sugere que precisa ser satisfeita a relação entre o número de observações e o número de variáveis respostas analisadas ($n - p$), evitando mau condicionamento das matrizes de dados a serem utilizadas. Sendo assim, subtraindo o número de observações (120) pelo número de variáveis (24), obtemos o valor de observações = 96, ou seja, $n > 50$.

Podemos observar que nas diferentes abordagens, para calcular o tamanho da amostra, o banco de dados mostrou-se satisfatório, podendo dar continuidade as outras premissas de utilização da análise fatorial.

3.1.2 Matriz de correlação

Uma matriz de correlação é usada mostrando as relações entre as variáveis individuais, ou seja, avalia a correlação entre duas variáveis. Hair Jr. *et al.* (2009), exemplifica que se inspeção visual não revela um número substancial de correlações maiores que 0,30, então a análise fatorial provavelmente é inadequada. Na Tabela 4, a maioria das correlações foram significativas, indicativo da adequacidade da utilização da análise fatorial.

Observaram-se correlações positivas significativas ($P < 0,05$) entre as características de peso corporal (PCV, PVA, PCQ e PCF) e medidas morfométricas (CEC, CIC, LG, LT, PG e PT). Como relatado por Rosa *et al.* (2014), medidas morfométricas são ferramentas que auxiliam na predição de características de interesse econômico, como peso e rendimento de

carcaça. Nesse contexto, podemos observar também que as características de rendimento de carcaça (RB, RCQ e RCF) apresentaram na sua maioria correlações significativas ($P < 0,05$) tanto para as características de peso corporal (PCV, PVA, PCQ e PCF) como para as medidas morfométricas.

Tabela 4. Correlação linear simples (abaixo da linha diagonal) e nível de probabilidade de significância (acima da linha diagonal) entre características de carcaças de ovinos da raça Santa Inês

	PCV	PVA	PCQ	PCF	RB	RCQ	RCF	CEC	CIC	LG	LT	PG	PP	PT	PER	LOM	COS	PES	PAL	RPER	RLOM	RCOS	RPES	RPAL	
PCV		*	*	*	*	ns	*	*	*	*	*	*	ns	*	*	*	*	*	*	*	*	ns	*	*	*
PVA	0.95		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	ns	*	*	*
PCQ	0.88	0.85		*	ns	*	*	*	*	*	*	*	ns	*	*	*	*	*	*	*	*	ns	*	*	*
PCF	0.86	0.83	0.99		ns	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	ns	*	*	*
RB	-0.38	-0.34	0.09	0.12		*	*	*	*	*	*	ns	*	ns	ns	ns	*	ns	ns	ns	ns	ns	ns	ns	ns
RCQ	-0.12	-0.28	0.26	0.28	0.78		*	ns	*	*	*	*	*	ns	ns	ns	*	*	*	ns	ns	*	ns	ns	ns
RCF	-0.18	-0.33	0.20	0.25	0.79	0.98		ns	*	*	*	ns	*	ns	ns	ns	*	*	*	ns	*	*	ns	ns	ns
CEC	0.58	0.55	0.53	0.52	-0.18	-0.05	-0.08		*	*	*	*	ns	*	*	*	ns	*	ns	*	ns	*	*	*	*
CIC	0.62	0.66	0.55	0.53	-0.23	-0.24	-0.28	0.63		*	*	*	*	*	*	*	ns	*	ns	ns	*	*	*	*	*
LG	0.65	0.63	0.54	0.51	-0.32	-0.19	-0.26	0.46	0.60		*	*	*	*	*	*	ns	*	ns	ns	*	*	*	*	*
LT	0.58	0.66	0.48	0.45	-0.31	-0.36	-0.41	0.34	0.63	0.83		*	*	*	*	*	ns	*	ns	ns	*	*	*	*	*
PG	0.46	0.34	0.50	0.50	0.00	0.28	0.25	0.30	0.28	0.44	0.35		*	*	*	*	*	ns	*	*	ns	ns	*	ns	ns
PP	-0.12	-0.14	-0.03	0.03	0.19	0.21	0.28	-0.15	-0.31	-0.30	-0.33	0.19		ns	ns	*	ns	*	ns	*	*	ns	ns	ns	ns
PT	0.40	0.39	0.42	0.42	-0.03	0.04	0.02	0.34	0.51	0.28	0.34	0.46	0.04		*	*	ns	*	ns	*	ns	*	*	*	*
PER	0.47	0.49	0.58	0.57	0.14	0.16	0.10	0.21	0.37	0.31	0.33	0.22	-0.13	0.19		*	*	*	*	*	*	*	ns	ns	ns
LOM	0.57	0.58	0.61	0.58	-0.02	0.05	-0.03	0.36	0.56	0.50	0.51	0.25	-0.31	0.28	0.72		*	*	*	*	*	*	ns	ns	ns
COS	0.39	0.28	0.56	0.56	0.26	0.50	0.46	0.11	0.10	0.12	-0.03	0.32	0.01	0.07	0.48	0.52		ns	*	ns	*	*	*	*	ns
PES	0.32	0.48	0.29	0.27	-0.12	-0.37	-0.40	0.24	0.57	0.41	0.61	-0.14	-0.29	0.27	0.41	0.60	-0.05		ns	*	*	*	*	*	*
PAL	0.33	0.30	0.49	0.50	0.25	0.36	0.33	0.07	0.05	0.08	0.02	0.32	0.05	0.03	0.56	0.52	0.63	0.03		ns	*	*	ns	*	*
RPER	-0.34	-0.28	-0.35	-0.38	0.01	-0.13	-0.16	-0.29	-0.10	-0.13	-0.05	-0.24	-0.18	-0.20	0.53	0.21	-0.07	0.19	0.08		*	*	*	*	*
RLOM	0.06	0.09	0.00	-0.05	-0.14	-0.17	-0.25	0.06	0.31	0.28	0.32	-0.05	-0.42	0.03	0.44	0.78	0.20	0.53	0.23	0.55		*	ns	*	*
RCOS	-0.34	-0.45	-0.27	-0.28	0.16	0.31	0.30	-0.34	-0.37	-0.32	-0.45	-0.07	-0.02	-0.32	-0.00	0.05	0.63	-0.34	0.24	0.26	0.27		ns	*	*
RPES	-0.46	-0.35	-0.42	-0.40	0.14	-0.12	-0.08	-0.33	-0.24	-0.39	-0.22	-0.36	0.14	-0.18	-0.14	-0.12	-0.21	0.36	0.01	0.22	0.14	0.10		ns	*
RPAL	-0.47	-0.49	-0.43	-0.44	0.13	0.10	0.10	-0.42	-0.45	-0.37	-0.40	-0.13	0.02	-0.36	0.00	-0.04	0.10	-0.24	0.55	0.42	0.27	0.51	0.38		*

ns = não significativo; * = significativo a 0,05 ($P < 0,05$); PCV = Peso de corpo vazio; PVA = Peso vivo ao abate; PCQ = Peso de carcaça quente; PCF = Peso de carcaça fria; RB = Rendimento biológico; RCQ = Rendimento de carcaça quente; RCF = Rendimento de carcaça fria; CEC = Comprimento externo da carcaça; CIC = Comprimento interno da carcaça; LG = Largura da garupa; LT = Largura do tórax; LG = Perímetro da garupa; PP = Perímetro do pernil; PT = Perímetro do tórax; PER = Pernel; LOM = Lombo; COS = Costela; PAL = Paleta; RPER = Rendimento do pernil; RLOM = Rendimento do lombo; RCOS = Rendimento da costela; RPES = Rendimento do pescoço; RPAL = Rendimento da paleta.

O mesmo ocorreu para as características de cortes cárneos (PER, LOM, COS, PAL), que apresentaram correlações significativas ($P < 0,05$) com suas respectivas características de rendimento (RPER, RLOM, RCOS, RPAL). Com descrito por Cezar e Sousa (2007a), as características de rendimento dos cortes cárneos em percentagem, é determinado por meio da relação entre o peso individual do corte e o peso total de todos os cortes da meia-carcaça reconstituída. Por outro lado, essas características apresentaram correlações não significativas ($P > 0,05$) com a maioria das características de peso corporal e medidas morfométricas. Esse resultado pode ter ocorrido em função das características de peso corporal e medidas morfométricas considerarem a carcaça animal como um todo, ocorrendo o inverso com os cortes cárneos, que se limitam a pontos específicos da carcaça animal.

3.1.3 Multicolinearidade

A análise fatorial lida com variáveis que geralmente são correlacionadas para reduzir a dimensão da análise para um número de fatores que não estão correlacionados (fatores independentes) (CONSTANTIN, 2014). Nesse sentido quando avaliamos um conjunto de variáveis altamente correlacionadas ocorre o fenômeno da multicolinearidade. De acordo com Yoo *et al.* (2014), multicolinearidade causam informações redundantes, o que significa que uma variável pode ser explicada pelas outras variáveis na análise, dificultando a interpretação dos resultados. Para a avaliação da multicolinearidade dos dados utilizam-se os testes de adequação da amostragem.

Na Tabela 5 são apresentados os resultados da adequação da amostra, com base no teste de esfericidade de Bartlett e o teste de KMO (Kaiser-Meyer-Olkin).

Tabela 5. Adequação da amostragem por meio do teste de esfericidade de Barlett e o teste de KMO (Kaiser-Meyer-Olkin).

	DF	276
Teste de esfericidade de Bartlett	Chi-Square	6482,92
	Pr > Chi Sq	<.0001
Teste KMO (Kaiser-Meyer-Olkin)	---	0,67

O valor encontrado para o teste de Bartlett foi estatisticamente significativo ($P < 0,05$). O teste de KMO apresentou valor de 0,67, uma vez que para proceder a análise fatorial, os dados analisados precisam ter um valor geral de $KMO > 0,50$. Com base nesses resultados, os dados de carcaças de ovinos Santa Inês demonstraram que a amostra utilizada é adequada.

3.1.4 Normalidade

Comumente, os testes estatísticos são baseados no pressuposto de que os dados coletados possuem distribuição normal (CAIN; ZHANG; YUAN, 2017). Para a realização de inferências na análise multivariada, segue-se o mesmo princípio da univariada. Para análise da normalidade multivariada, em geral, os testes existentes são extensões dos propostos para a análise univariada. Exemplo disso é o teste de Mardia, que é uma extensão do teste univariado de Shapiro-Wilk, que permite avaliar a normalidade multivariada. Na Tabela 6 é apresentado o teste de normalidade multivariada Mardia, que baseia em assimetria e curtose.

Tabela 6. Teste de normalidade multivariada Mardia para as variáveis de carcaças ovinas da raça Santa Inês.

Teste estatístico	Valor	Probabilidade
Mardia assimetria	6058	<.0001
Mardia curtose	20.62	<.0001

Como podemos observar, as variáveis de características de carcaças de ovinos da raça Santa Inês não atenderam a premissa de normalidade multivariada ($P < 0,05$). Todavia, Hair Jr. *et al.* (2009) descreve que do ponto de vista estatístico, a aplicação da premissa de normalidade aplicam-se apenas porque a não normalidade diminuem as correlações observadas. Na Tabela 4 podemos observar que a maioria das variáveis apresentaram correlações significativas ($P < 0,05$) umas com as outras, sendo suficiente para produzir fatores representativos na análise fatorial. Da mesma forma, Williams, Onsmann e Brown (2010) e Howard (2016) publicaram revisões objetivando fornecer protocolos aos pesquisadores para a aplicação da análise fatorial e destacaram como premissas principais o tamanho da amostra, a matriz de correlações e a adequação da amostragem pelo teste de Bartlett e KMO. E nesse estudo foi possível observar que todas essas premissas foram atendidas.

3.1.5 Retenção e escolha do número de fatores

Na Tabela 7 são apresentados os autovalores da matriz de correlação das 24 variáveis analisadas. Podemos observar que retenção de fatores mediante análise de componentes principais e a escolha do número de fatores pelo critério de Kaiser foi possível extrair seis fatores (autovalores > 1).

Tabela 7. Autovalores da matriz de correlação das 24 variáveis analisadas pelo critério de Kaiser

	Autovalores	Diferença	Proporção	Cumulativo
1	8.09095426	3.66653279	0.3371	0.3371
2	4.42442147	0.99411106	0.1844	0.5215
3	3.43031041	1.88116695	0.1429	0.6644
4	1.54914346	0.29800618	0.0645	0.7290
5	1.25113728	0.20110289	0.0521	0.7811
6	1.05003439	0.13744193	0.0438	0.8248

Esses seis fatores foram capazes de explicar cerca de 82,48% da variância original das variáveis. Esse valor demonstrou-se bastante expressivo, representando pouca perda de informações, indicando que a análise resumiu a maior parte da informação em um número mínimo de fatores necessários para explicar o máximo de variância representada pelas variáveis originais.

3.1.6 Rotações dos Fatores

Na Tabela 8 são apresentadas as cargas fatoriais dos seis fatores retidas pelo método de Kaiser, sem a utilização de rotações dos fatores, assim como as comunalidades. A inserção dessa tabela tem como finalidade observar o comportamento dos fatores não rotacionados e os mesmos fatores obtidos com diferentes tipos de rotações. A não utilização de rotação representa que a variação de cada fator será igual a seu autovalor ou valor próprio.

Tabela 8. Cargas fatoriais dos seis fatores sem a utilização de rotação dos fatores e suas respectivas comunalidades.

Variáveis	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Comunalidade
PCV	0.920	0.022	-0.150	-0.231	0.191	-0.107	0,97
PVA	0.923	-0.108	-0.084	-0.088	0.273	-0.161	0,98
PCQ	0.881	0.379	-0.160	0.034	0.120	-0.155	0,98
PCF	0.858	0.404	-0.197	0.057	0.149	-0.158	0,99
RB	-0.219	0.688	-0.016	0.564	-0.179	-0.088	0,88
RCQ	-0.092	0.895	-0.145	0.222	-0.269	-0.001	0,95
RCF	-0.157	0.886	-0.196	0.247	-0.215	-0.002	0,95
CEC	0.648	-0.063	-0.213	0.003	-0.146	-0.089	0,50
CIC	0.789	-0.238	0.054	0.150	-0.172	0.035	0,73
LG	0.768	-0.206	0.016	-0.136	-0.211	0.168	0,72
LT	0.751	-0.369	0.100	0.063	-0.104	0.181	0,76
PG	0.490	0.356	-0.255	-0.178	-0.036	0.635	0,87
PP	-0.229	0.290	-0.386	0.141	0.535	0.310	0,69
PT	0.520	0.016	-0.212	0.210	-0.075	0.459	0,62
PER	0.581	0.329	0.521	0.136	0.072	0.026	0,74
LOM	0.731	0.187	0.576	0.054	-0.071	-0.029	0,91
COS	0.355	0.744	0.238	-0.296	-0.002	-0.193	0,86
PES	0.526	-0.382	0.444	0.509	0.120	-0.137	0,91
PAL	0.316	0.644	0.387	-0.150	0.415	0.059	0,86
RPER	-0.207	-0.080	0.769	0.087	-0.101	0.214	0,70
RLOM	0.261	-0.108	0.850	0.000	-0.223	0.105	0,86
RCOS	-0.383	0.476	0.431	-0.431	-0.165	-0.078	0,78
RPES	-0.413	-0.135	0.368	0.480	0.440	-0.032	0,75
RPAL	-0.492	0.274	0.564	-0.234	0.277	0.227	0,82

Peso de corpo vazio (PCV); Peso vivo ao abate (PVA); Peso de carcaça quente (PCQ); Peso de carcaça fria (PCF); Rendimento biológico (RB); Rendimento de carcaça quente (RCQ); Rendimento de carcaça fria (RCF); Comprimento externo da carcaça (CEC); Comprimento interno da carcaça (CIC); Largura da garupa (LG); Largura do tórax (LT); Perímetro da garupa (PG); Perímetro do pernil (PP); Perímetro do tórax (PT); Pernil (PER); Lombo (LOM); Costela (COS); Pescoço (PES); Paleta (PAL); Rendimento do pernil (RPER); Rendimento do lombo (RLOM); Rendimento da costela (RCOS); Rendimento do pescoço (RPES); Rendimento da paleta (RPAL).

Para fins de interpretação dos fatores temos que levar em consideração as cargas fatoriais que são significativas. Neste estudo, as cargas fatoriais foram consideradas significativas em função do tamanho da amostra (carga fatorial $\geq 0,50$), sendo que as cargas fatoriais representam a correlação entre uma variável original e seu fator.

Na Tabela 8 podemos observar a presença de cargas fatoriais cruzadas para as variáveis RB, PER, LOM e PES, ou seja, existem cargas fatoriais significativas ($\geq 0,50$) na mesma linha da variável, o que dificulta a interpretação dos fatores, uma vez que devemos considerar um fator independente do outro. Também se observou que a variável RPES não apresentou cargas fatoriais significativas ($< 0,50$) em nenhum dos seis fatores, o que sugere a necessidade de rotação dos fatores para melhor interpretabilidade.

Williams, Onsmann e Brown (2010) descreveram que as técnicas de rotações maximizam as altas cargas fatoriais e minimiza as baixas cargas fatoriais, produzindo, assim, uma interpretação mais simplificada dos dados. Assim sendo, a rotação altera a distribuição da proporção da variação explicada por cada fator, no entanto, a variação total explicada por todos os fatores permanece a mesma já obtida na análise sem rotação (82%) (Tabela 7).

A comunalidade apresentada na Tabela 8 é a proporção de variabilidade de cada variável que é explicada pelos fatores. Quanto mais próximo de 1, melhor a variável explica a variação nos fatores. Hair Jr. *et al.* (2009) descrevem que pelo menos metade da variância de cada variável deve ser levada em conta; usando essa diretriz, todas as variáveis com comunalidade menores que 0,50 apresentam explicação insuficiente. Na Tabela 8, a variável que apresentou menor comunalidade foi a CEC (0,50), mesmo assim, obteve a proporção da variabilidade adequada para explicar os fatores.

3.1.6.1 Rotações oblíquas

Nas Tabelas 9 e 10 são apresentadas as correlações entre os fatores obtidos pela rotação oblíqua Oblimin direta e as cargas fatoriais dos fatores, respectivamente. O Oblimin direto foi desenvolvido para a complexidade variável, age minimizando a complexidade variável/linha (SASS; SCHMITT, 2010).

Tabela 9. Correlações entre os fatores, por meio da rotação oblíqua Oblimin direta.

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Factor1	1.00000					
Factor2	-0.14057	1.00000				
Factor3	-0.16992	-0.01904	1.00000			
Factor4	-0.40390	0.07583	0.27396	1.00000		
Factor5	0.17549	-0.25688	0.01025	0.14390	1.00000	
Factor6	0.06323	0.17599	-0.07807	-0.11497	0.08555	1.00000

Rotações oblíquas sugerem a existência de correlações entre fatores. Tabachnick e Fidell (2018) recomendam que se tivermos apenas algumas correlações acima de 0,30, podemos concluir que não há estrutura subjacente entre os fatores. Observa-se na Tabela 9 que apenas o Fator1 e o Fator4 apresentaram correlação acima de 0,30, indicando que não existe uma estrutura de correlação subjacente entre os fatores extraídos.

Também se verificaram cargas fatoriais cruzadas em 10 variáveis (PCV, PCQ, PVA, PCF, CIC, LG, LT, PAL, RCOS e RPAL) (Tabela 10). Comparando esse resultado com o da Tabela 8 (sem utilização de rotação), podemos concluir que as cargas fatoriais sem rotação

apresentaram melhores resultados, uma vez que apresentaram menor número de cargas fatoriais cruzadas (quatro). Esse comparativo é muito importante, indicando que nem sempre a escolha de uma determinada rotação contribuirá para reduzir uma função de complexidade e propor uma estrutura mais simples e de fácil interpretação.

Tabela 10. Cargas fatoriais dos seis fatores, obtidas pela rotação oblíqua Oblimin direta e suas respectivas comunalidades.

Variáveis	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Comunalidade
PCV	0.832	-0.338	-0.266	0.601	-0.413	0.062	0,97
PCQ	0.913	0.095	-0.249	0.596	-0.328	0.086	0,98
PVA	0.797	-0.417	-0.242	0.564	-0.420	0.274	0,98
PCF	0.904	0.127	-0.284	0.588	-0.284	0.087	0,99
RB	0.019	0.903	0.054	-0.102	0.219	0.042	0,88
RCQ	0.204	0.943	-0.021	0.039	0.181	-0.348	0,95
RCF	0.148	0.953	-0.070	0.003	0.256	-0.337	0,95
CEC	0.445	-0.172	-0.285	0.504	-0.494	0.125	0,50
CIC	0.476	-0.322	-0.038	0.627	-0.623	0.384	0,73
LG	0.474	-0.391	-0.026	0.663	-0.591	0.105	0,72
LT	0.397	-0.482	0.024	0.649	-0.557	0.371	0,76
PG	0.421	0.104	-0.109	0.783	0.045	-0.327	0,87
PP	-0.063	0.251	-0.297	0.080	0.729	-0.070	0,69
PT	0.256	-0.002	-0.160	0.763	-0.151	0.234	0,62
PER	0.699	0.125	0.469	0.304	-0.204	0.213	0,74
LOM	0.752	-0.034	0.495	0.371	-0.458	0.234	0,91
COS	0.733	0.425	0.239	0.039	-0.059	-0.414	0,86
PES	0.313	-0.326	0.272	0.238	-0.403	0.836	0,91
PAL	0.719	0.288	0.391	0.085	0.288	-0.171	0,86
RPER	-0.154	-0.020	0.803	-0.225	0.028	0.151	0,70
RLOM	0.229	-0.173	0.818	0.033	-0.373	0.212	0,86
RCOS	0.002	0.359	0.517	-0.493	0.174	-0.594	0,78
RPES	-0.290	0.049	0.320	-0.412	0.461	0.518	0,75
RPAL	-0.128	0.164	0.656	-0.460	0.567	-0.285	0,82

Peso de corpo vazio (PCV); Peso vivo ao abate (PVA); Peso de carcaça quente (PCQ); Peso de carcaça fria (PCF); Rendimento biológico (RB); Rendimento de carcaça quente (RCQ); Rendimento de carcaça fria (RCF); Comprimento externo da carcaça (CEC); Comprimento interno da carcaça (CIC); Largura da garupa (LG); Largura do tórax (LT); Perímetro da garupa (PG); Perímetro do pernil (PP); Perímetro do tórax (PT); Pernil (PER); Lombo (LOM); Costela (COS); Pescoço (PES); Paleta (PAL); Rendimento do pernil (RPER); Rendimento do lombo (RLOM); Rendimento da costela (RCOS); Rendimento do pescoço (RPES); Rendimento da paleta (RPAL).

De acordo com Sass e Schmitt (2010), a rotação oblíqua Oblimin direta foi desenvolvida com o objetivo de aperfeiçoar apenas a estrutura simples e garantir que grupos de variáveis sejam independentes e facilmente interpretáveis. No entanto, muitas variáveis não são medidas perfeitas para cada fator e, portanto, podem resultar em complexidade fatorial (cargas cruzadas). Com base no exposto, o método oblíquo Oblimin direta parece funcionar bem

quando existem diferentes grupos, o que não é o caso deste estudo, uma vez que as cargas fatoriais se apresentaram cruzadas. Para Hair Jr. *et al.* (2009), o pesquisador pode descobrir que diferentes métodos de rotações eliminam cargas cruzadas. Porém, se uma variável persiste em apresentar cargas cruzadas, ela torna-se candidata à eliminação.

Nas Tabelas 11 e 12 são apresentadas as correlações entre fatores por meio da rotação oblíqua Promax e as cargas fatoriais dos fatores, respectivamente. Na Tabela 11 observa-se que dos seis fatores extraídos, dois apresentaram correlações acima de 0,30 (Fator1 com o Fator5 e o Fator2 com o Fator4), sendo que este último não foi observado na rotação oblíqua Oblimin direta (Tabela 9).

Tabela 11. Correlações entre os fatores, por meio do método de rotação oblíqua Promax.

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Factor1	1.00000					
Factor2	-0.02258	1.00000				
Factor3	-0.24675	0.24944	1.00000			
Factor4	0.02320	-0.38953	-0.12812	1.00000		
Factor5	-0.40916	0.01192	0.04163	-0.02589	1.00000	
Factor6	0.09440	0.05320	-0.14596	0.07950	-0.11557	1.00000

Outro ponto em destaque comparando ambas as rotações oblíquas é que a rotação Oblimin direta apresentou correlação acima de 0,30 entre o Fator1 e o Fator4 (Tabela 9). Na rotação Promax a correlação apresentou muito baixa (0,02) (Tabela 11), no entanto, as cargas fatoriais que eram significativas no Fator4 (Tabela 10) apresentaram significativas (exceto, PG) no Fator1 (Tabela 12). Nesse sentido, a rotação oblíqua Promax não proporcionou correlações entre Fator1 e Fator4, porém permitiu correlação intrafatorial com as cargas fatoriais que apresentaram correlações significativas na rotação oblíqua Oblimin direta. Isso ocorreu provavelmente pelo fato de a rotação oblíqua Promax ser baseada na rotação ortogonal Varimax, fazendo com que a rotação oblíqua Promax contribua com mais peso na complexidade da coluna/fator, diferindo do Oblimin direta que coloca mais peso na complexidade da linha/variável.

Nas Tabelas 9, 10, 11 e 12 são apresentadas as correlações entre os fatores obtidos pelas rotações oblíquas Oblimin direta e Promax e as cargas fatoriais dos fatores, respectivamente. O Oblimin direto foi desenvolvido para a complexidade variável, age minimizando a complexidade variável/linha (SASS; SCHMITT, 2010). Em contrapartida, a rotação oblíqua Promax é baseada na rotação ortogonal Varimax, age minimizando a complexidade da coluna/fator.

Tabela 12. Cargas fatoriais dos seis fatores, obtidas pela rotação oblíqua Promax e suas respectivas comunalidades.

Variáveis	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Comunalidade
PCV	0.931	-0.201	0.181	0.138	0.385	-0.130	0,97
PCQ	0.929	0.237	0.123	0.290	0.389	-0.073	0,98
PVA	0.940	-0.283	0.250	0.047	0.349	0.069	0,98
PCF	0.918	0.268	0.070	0.295	0.389	-0.056	0,99
RB	-0.146	0.893	-0.151	0.254	-0.057	0.130	0,88
RCQ	-0.030	0.956	-0.249	0.447	0.057	-0.258	0,95
RCF	-0.080	0.959	-0.331	0.426	0.039	-0.219	0,95
CEC	0.666	-0.061	0.175	-0.163	0.286	-0.142	0,50
CIC	0.734	-0.215	0.516	-0.228	0.421	0.069	0,73
LG	0.683	-0.292	0.458	-0.135	0.481	-0.173	0,72
LT	0.660	-0.395	0.538	-0.250	0.494	0.097	0,76
PG	0.418	0.155	-0.060	0.267	0.811	-0.299	0,87
PP	-0.151	0.216	-0.634	0.211	0.277	0.242	0,69
PT	0.477	0.060	0.124	-0.151	0.725	0.115	0,62
PER	0.491	0.176	0.576	0.482	0.270	0.191	0,74
LOM	0.616	0.037	0.757	0.377	0.262	0.096	0,91
COS	0.347	0.477	0.131	0.770	-0.017	-0.352	0,86
PES	0.472	-0.276	0.675	-0.207	0.137	0.643	0,91
PAL	0.272	0.301	0.133	0.871	0.180	0.068	0,86
RPER	-0.373	-0.105	0.587	0.198	-0.067	0.254	0,70
RLOM	0.068	-0.190	0.890	0.212	0.045	0.137	0,86
RCOS	-0.446	0.291	0.073	0.601	-0.383	-0.400	0,78
RPES	-0.410	-0.037	0.039	0.032	-0.214	0.746	0,75
RPAL	-0.595	0.043	0.054	0.609	-0.173	0.097	0,82

Peso de corpo vazio (PCV); Peso vivo ao abate (PVA); Peso de carcaça quente (PCQ); Peso de carcaça fria (PCF); Rendimento biológico (RB); Rendimento de carcaça quente (RCQ); Rendimento de carcaça fria (RCF); Comprimento externo da carcaça (CEC); Comprimento interno da carcaça (CIC); Largura da garupa (LG); Largura do tórax (LT); Perímetro da garupa (PG); Perímetro do pernil (PP); Perímetro do tórax (PT); Pernil (PER); Lombo (LOM); Costela (COS); Pescoço (PES); Paleta (PAL); Rendimento do pernil (RPER); Rendimento do lombo (RLOM); Rendimento da costela (RCOS); Rendimento do pescoço (RPES); Rendimento da paleta (RPAL).

Diferente das outras rotações oblíquas que se desenvolvem em um único estágio para permitir a correlações entre os fatores, a rotação oblíqua Promax é considerada um método de rotação indireta. Ele primeiro executa uma rotação ortogonal para posteriormente executar a rotação oblíqua. A rotação oblíqua Promax é baseada na rotação ortogonal Varimax, a diferença é que esta permite correlações entre os fatores. Para que isso ocorra, elavam-se as cargas fatoriais da rotação ortogonal Varimax a alguma potência especificada (geralmente 4), girando a matriz resultante, permitindo a correlação entre os fatores (FINCH, 2011).

De acordo com Williams, Onsmann e Brown (2010), independentemente de qual método de rotação é usado, deve-se sempre escolher aquele que produz o melhor ajuste e adequação fatorial, tanto intuitivamente quanto conceitualmente. Partindo desse pressuposto, as rotações

oblíquas Oblimin direta e Promax, dos seis fatores extraídos apresentaram um e dois fatores correlacionados acima de 0,30 (Tabela 9 e 10, respectivamente), sendo que as correlações entre os demais fatores foram muito baixas, a maioria inferior a 0,20. Krabbe (2016) sugere que nestas condições é razoável considerar as rotações ortogonais e uma rotação oblíqua deve ser usada apenas se houver boas razões para supor que os fatores subjacentes possam estar correlacionados. Neste estudo, essa recíproca não é verdadeira, uma vez que as rotações oblíquas apresentaram apenas dois fatores correlacionados dos seis fatores extraídos, por isso devemos optar pelas rotações ortogonais.

Do ponto de vista das cargas fatoriais, a rotação oblíqua Oblimin direta (Tabela 10) foi menos eficiente comparado à rotação oblíqua Promax (Tabela 12). A rotação Oblimin direta apresentou 10 variáveis cruzadas (PCV, PCQ, PVA, PCF, CEC, CIC, LT LG, RCOS, RPAL) em comparação à rotação oblíqua Promax, que apresentou cinco variáveis cruzadas (CIC, LT, LOM, PES, RPAL). Por outro lado, quando comparamos as rotações oblíquas (Oblimin direta e Promax) com os fatores extraídos sem retenção, esse último demonstrou mais eficiente, uma vez que apresentou número inferior de variáveis cruzadas, quatro no total (Tabela 8), reforçando mais uma vez que devemos optar pelas rotações ortogonais.

3.1.6.2 Rotações ortogonais

Na Tabela 13 são apresentadas as cargas fatoriais obtidas pela rotação ortogonal Quartimax, que foi o primeiro método criado da rotação ortogonal (HOWARD, 2016). Esse método, assim como a Oblimin direta, é projetado para minimizar a complexidade das cargas fatoriais pelo indicador variável (FINCH, 2011), ou seja, coloca mais peso na complexidade linha/variável, diferenciando apenas ao fato de não considerar a correlação entre os fatores.

Tabela 13. Cargas fatorais dos seis fatores retidos pelo método de Kaiser, obtido pela rotação ortogonal Quartmax e suas respectivas comunalidades.

Variáveis	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Comunalidade
PCV	0.911	-0.188	-0.134	0.275	-0.090	-0.009	0,97
PVA	0.921	-0.253	-0.088	0.207	0.101	-0.068	0,98
PCQ	0.895	0.253	-0.131	0.317	0.007	-0.012	0,98
PCF	0.879	0.279	-0.179	0.320	0.032	-0.005	0,99
RB	-0.169	0.899	0.014	0.018	0.202	-0.012	0,88
RCQ	-0.060	0.927	-0.093	0.206	-0.167	0.092	0,95
RCF	-0.115	0.923	-0.165	0.191	-0.122	0.099	0,95
CEC	0.682	-0.002	-0.045	-0.120	-0.127	-0.029	0,50
CIC	0.780	-0.110	0.273	-0.188	0.030	0.054	0,73
LG	0.731	-0.216	0.235	-0.083	-0.222	0.170	0,72
LT	0.716	-0.295	0.305	-0.180	0.030	0.174	0,76
PG	0.439	0.149	-0.102	0.202	-0.300	0.714	0,87
PP	-0.202	0.139	-0.554	0.208	0.309	0.425	0,69
PT	0.522	0.135	0.022	-0.206	0.093	0.527	0,62
PER	0.486	0.203	0.487	0.435	0.185	0.044	0,74
LOM	0.630	0.091	0.611	0.353	0.067	-0.037	0,91
COS	0.293	0.412	0.110	0.704	-0.277	-0.135	0,86
PES	0.500	-0.165	0.488	-0.150	0.590	-0.155	0,91
PAL	0.209	0.221	0.129	0.850	0.122	0.122	0,86
RPER	-0.333	-0.080	0.736	0.121	0.151	0.075	0,70
RLOM	0.119	-0.129	0.894	0.168	0.029	-0.027	0,86
RCOS	-0.472	0.203	0.277	0.506	-0.396	-0.154	0,78
RPES	-0.431	-0.042	0.150	0.037	0.730	-0.059	0,75
RPAL	-0.622	-0.051	0.287	0.565	0.067	0.146	0,82

Peso de corpo vazio (PCV); Peso vivo ao abate (PVA); Peso de carcaça quente (PCQ); Peso de carcaça fria (PCF); Rendimento biológico (RB); Rendimento de carcaça quente (RCQ); Rendimento de carcaça fria (RCF); Comprimento externo da carcaça (CEC); Comprimento interno da carcaça (CIC); Largura da garupa (LG); Largura do tórax (LT); Perímetro da garupa (PG); Perímetro do pernil (PP); Perímetro do tórax (PT); Pernil (PER); Lombo (LOM); Costela (COS); Pescoço (PES); Paleta (PAL); Rendimento do pernil (RPER); Rendimento do lombo (RLOM); Rendimento da costela (RCOS); Rendimento do pescoço (RPES); Rendimento da paleta (RPAL).

A rotação ortogonal Quartimax, apesar de apresentar critérios de rotações semelhantes à rotação oblíqua Oblimin direta, por colocar mais peso na linha/variável, apresentou comportamento comparável à rotação oblíqua Promax (Tabela 12), que coloca mais peso no fator/coluna. A justificativa para esse resultado pode estar diretamente relacionada ao fato de rotação ortogonal Quartimax não considerar correlações entre os fatores, interpretando os fatores de forma independente. Nesse sentido, a rotação ortogonal Quartimax promoveu a inserção das variáveis em um único fator (Fator1), ou seja, as variáveis que estava no Fator 4 migraram para o Fator1, assim como fez a rotação oblíqua Promax. Na rotação oblíqua Promax (Tabela 12) o Fator1 e o Fator4 não foram correlacionados, mas as variáveis do Fator4 migraram para o Fator1, demonstrando correlações entre as variáveis. De forma indireta, isso

promove o mesmo efeito da rotação oblíqua Oblimin direta (Tabela 10), que apesar de estar em fatores distintos (1 e 4), mas por ser oblíqua, permite a correlação entre os fatores, ou seja, as variáveis também estão correlacionadas. De acordo com Osborne (2015), nas rotações oblíquas, os eixos podem assumir qualquer posição no espaço fatorial e que dois fatores altamente correlacionados podem ser melhor interpretados como apenas um fator.

Quanto às variáveis cruzadas, a rotação ortogonal Quartmax apresentou melhor desempenho comparado às rotações oblíquas Oblimin direta e Promax. As rotações oblíquas Oblimin direta (Tabela 10) e Promax (Tabela 12) apresentaram 10 e 5 variáveis cruzadas, respectivamente. A rotação ortogonal Quartmax apresentou quatro variáveis cruzadas (PT, LOM, PES e RPAL), obtendo a mesma quantidade de cargas cruzadas que a fatores não rotacionados (Tabela 8).

Na Tabela 14 são apresentadas as cargas fatoriais obtidas pela rotação ortogonal Equamax. A rotação ortogonal Equamax representa a combinação dos critérios das rotações ortogonais Quartimax e Varimax (FINCH, 2011). Em outras palavras, a rotação ortogonal Equamax coloca ênfase tanto na complexidade da linha/variável como na complexidade da coluna/fator. Portanto, essa relação precisa ser considerada para entender se a rotação Equamax visa mais a complexidade de variáveis ou fatores e, nesse sentido, os pesquisadores precisam estar cientes de que a função de cada critério de rotação pode mudar com base em p (número de variáveis) e m (número de fatores) (SASS; SCHMITT, 2010).

Tabela 14. Cargas fatoriais dos seis fatores obtidas pela rotação ortogonal Equamax e suas respectivas comunalidades.

Variáveis	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Comunalidade
PCV	0.723	-0.266	-0.025	-0.261	-0.424	0.352	0,97
PVA	0.716	-0.333	0.036	-0.420	-0.322	0.267	0,98
PCQ	0.785	0.175	-0.002	-0.310	-0.350	0.343	0,98
PCF	0.790	0.202	-0.048	-0.319	-0.319	0.340	0,99
RB	-0.002	0.909	0.035	-0.071	0.198	-0.081	0,88
RCQ	0.127	0.932	-0.086	0.205	-0.040	0.124	0,95
RCF	0.103	0.932	-0.158	0.194	0.022	0.101	0,95
CEC	0.289	-0.063	-0.000	-0.357	-0.471	0.246	0,50
CIC	0.248	-0.182	0.330	-0.509	-0.405	0.327	0,73
LG	0.226	-0.280	0.265	-0.257	-0.497	0.457	0,72
LT	0.175	-0.360	0.350	-0.448	-0.326	0.407	0,76
PG	0.204	0.115	-0.085	0.116	-0.149	0.876	0,87
PP	0.093	0.159	-0.517	0.035	0.556	0.273	0,69
PT	0.040	0.087	0.055	-0.411	-0.077	0.658	0,62
PER	0.558	0.157	0.594	-0.082	0.003	0.213	0,74
LOM	0.550	0.033	0.709	-0.128	-0.208	0.205	0,91
COS	0.677	0.389	0.184	0.402	-0.220	0.069	0,86
PES	0.230	-0.217	0.585	-0.676	0.084	-0.061	0,91
PAL	0.739	0.205	0.245	0.326	0.254	0.206	0,86
RPER	-0.211	-0.054	0.716	0.217	0.298	-0.081	0,70
RLOM	0.055	-0.143	0.910	0.085	-0.042	0.015	0,86
RCOS	0.038	0.249	0.235	0.779	0.000	-0.229	0,78
RPES	-0.097	-0.009	0.192	-0.181	0.745	-0.336	0,75
RPAL	-0.002	0.006	0.277	0.655	0.549	-0.094	0,82

Peso de corpo vazio (PCV); Peso vivo ao abate (PVA); Peso de carcaça quente (PCQ); Peso de carcaça fria (PCF); Rendimento biológico (RB); Rendimento de carcaça quente (RCQ); Rendimento de carcaça fria (RCF); Comprimento externo da carcaça (CEC); Comprimento interno da carcaça (CIC); Largura da garupa (LG); Largura do tórax (LT); Perímetro da garupa (PG); Perímetro do pernil (PP); Perímetro do tórax (PT); Pernil (PER); Lombo (LOM); Costela (COS); Pescoço (PES); Paleta (PAL); Rendimento do pernil (RPER); Rendimento do lombo (RLOM); Rendimento da costela (RCOS); Rendimento do pescoço (RPES); Rendimento da paleta (RPAL).

A rotação ortogonal Equamax apresentou quatro variáveis cruzadas (PER, LOM, PES, RPAL), apresentando um desempenho inferior às suas idealizadoras (rotações ortogonais Quartmax e Varimax). As rotações ortogonais Quartmax e Varimax apresentaram quatro (Tabela 10) e três (Tabela 15) variáveis cruzadas, respectivamente.

A rotação ortogonal Equamax apresentou quatro variáveis cruzadas (PER, LOM, PES, RPAL), apresentando desempenho semelhante a rotação Quartmax, que também apresentou quatro variáveis cruzadas.

Como descrito por Sass e Schmitt (2010), a rotação ortogonal Equamax minimiza a complexidade de variáveis e fatores com base nas variáveis p, em que o p maior minimiza a complexidade das variáveis e o p menor minimiza a complexidade dos fatores. Nesse estudo,

ao que parece, a rotação ortogonal Equamax agiu minimizando a complexidade da coluna/fator, comportando como a rotação ortogonal Varimax, que coloca mais peso na complexidade coluna/fator. Porém, essa rotação gerou uma problemática por priorizar a complexidade coluna/fator, tendo como consequência o aumento da complexidade da linha/variável. Como pode ser visto na Tabela 14, as variáveis LG e LT não apresentaram cargas fatoriais significativas ($< 0,50$), dificultando a interpretação da análise.

Geralmente quando o objetivo é minimizar a complexidade linha/variável, a análise é mais sensível a cargas cruzadas. Na rotação ortogonal Equamax houve tanto a problemática das cargas fatoriais cruzadas, além do “excesso” de peso nas cargas fatoriais coluna/fator que acarretou a redução das cargas fatoriais linha/variável, chegando ao ponto das variáveis LG e LT não apresentarem cargas fatoriais significativas; anteriormente isso foi observado somente nos fatores extraídos sem rotação (Tabela 8).

Na Tabela 15 são apresentadas as cargas fatoriais obtidas pela rotação ortogonal Varimax. O objetivo da rotação da rotação ortogonal Varimax é criar uma estrutura simples maximizando as diferenças entre cargas dentro dos fatores e entre variáveis (FINCH, 2011), ou seja, busca minimizar o número de variáveis que têm altas cargas em um fator.

Tabela 15. Cargas fatoriais dos seis fatores obtidos pela rotação ortogonal Varimax e suas respectivas comunalidades.

Variável	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Comunalidade
PCV	0.908	-0.228	0.013	0.212	0.158	-0.140	0,97
PVA	0.925	-0.290	0.066	0.140	0.105	0.050	0,98
PCQ	0.919	0.215	0.019	0.251	0.164	-0.045	0,98
PCF	0.914	0.242	-0.031	0.252	0.165	-0.021	0,99
RB	-0.114	0.908	-0.002	0.019	-0.019	0.203	0,88
RCQ	-0.017	0.927	-0.104	0.206	0.092	-0.170	0,95
RCF	-0.056	0.926	-0.184	0.191	0.083	-0.123	0,95
CEC	0.646	-0.035	0.085	-0.162	0.136	-0.166	0,50
CIC	0.663	-0.152	0.405	-0.206	0.253	-0.002	0,73
LG	0.590	-0.260	0.344	-0.087	0.344	-0.248	0,72
LT	0.565	-0.338	0.410	-0.179	0.349	0.007	0,76
PG	0.297	0.108	-0.096	0.243	0.778	-0.303	0,87
PP	-0.133	0.145	-0.629	0.220	0.319	0.319	0,69
PT	0.378	0.094	0.073	-0.180	0.651	0.080	0,62
PER	0.422	0.178	0.539	0.438	0.143	0.170	0,74
LOM	0.533	0.059	0.696	0.350	0.107	0.046	0,91
COS	0.347	0.403	0.149	0.675	-0.101	-0.296	0,86
PES	0.435	-0.186	0.579	-0.164	-0.017	0.569	0,91
PAL	0.242	0.213	0.113	0.849	0.102	0.117	0,86
RPER	-0.454	-0.072	0.646	0.196	0.017	0.189	0,70
RLOM	-0.035	-0.142	0.888	0.218	0.025	0.044	0,86
RCOS	-0.442	0.228	0.187	0.533	-0.271	-0.369	0,78
RPES	-0.380	-0.016	0.067	0.059	-0.165	0.754	0,75
RPAL	-0.624	-0.024	0.129	0.629	-0.038	0.113	0,82

Peso de corpo vazio (PCV); Peso vivo ao abate (PVA); Peso de carcaça quente (PCQ); Peso de carcaça fria (PCF); Rendimento biológico (RB); Rendimento de carcaça quente (RCQ); Rendimento de carcaça fria (RCF); Comprimento externo da carcaça (CEC); Comprimento interno da carcaça (CIC); Largura da garupa (LG); Largura do tórax (LT); Perímetro da garupa (PG); Perímetro do pernil (PP); Perímetro do tórax (PT); Pernil (PER); Lombo (LOM); Costela (COS); Pescoço (PES); Paleta (PAL); Rendimento do pernil (RPER); Rendimento do lombo (RLOM); Rendimento da costela (RCOS); Rendimento do pescoço (RPES); Rendimento da paleta (RPAL).

Com base nas cargas cruzadas podemos observar que a rotação ortogonal Varimax (Tabela 15) foi a que apresentou melhor resultado com apenas três variáveis cruzadas (LOM, PES e RPAL), comparando com as demais rotações, tanto oblíquas como ortogonais. Como descrito por Howard (2016), a escolha da rotação deve ser baseada em uma teoria a priori, nesse caso, se o pesquisador acredita que os fatores resultantes não estão correlacionados deve executar uma rotação ortogonal; por outro lado, se o pesquisador acredita que os fatores resultantes estão correlacionados, deve executar uma rotação oblíqua. Por segurança, convém sempre testar todas as possibilidades para uma tomada de decisão mais consciente.

3.1.6.3 Interpretação dos fatores

Dentre as possibilidades para a tomada de decisão sobre qual rotação será utilizada para a interpretação dos fatores, a decisão mais consciente é optar pela rotação ortogonal Varimax, por apresentar número inferior de variáveis cruzadas e por não apresentar cargas fatoriais não significativas ($< 0,50$). Treiblmaier e Filzmoser (2010) relataram que a rotação ortogonal produz resultados mais estáveis, quando os fatores não são realmente correlacionados.

A interpretação dos fatores da análise fatorial exige do pesquisador examinar quais variáveis são atribuíveis a um fator, permitindo nomear o fator, nomeando-o. Para isso, duas ou três variáveis devem ter forte associação com um fator para que possa ser atribuída uma interpretação significativa (WILLIAMS; ONSMAN; BROWN, 2010).

No Fator1 (Tabela 15), as maiores cargas foram obtidas para as variáveis peso corporal e medidas morfométricas, o que nos permite nomear esse fator como “Fator Capacidade Corporal”. A variável LOM não será considerada no Fator1, em virtude de ser uma variável cruzada com o Fator3. Esse resultado do Fator1 é pertinente com o que observamos na Tabela 4, em que as características de peso corporal e medidas morfométricas apresentaram correlações significativas ($P < 0,05$).

O Fator2 apresentou cargas fatoriais significativas para as características de rendimentos, sendo então denominada fator “Rendimento”. Apesar das correlações significativas entre peso corporal e características de rendimento (Tabela 4), as variáveis RB, RCQ e RCF foram altamente correlacionadas ($r = 0,78, 0,79$ e $0,98$), respectivamente, o que contribuiu para as cargas fatoriais maiores no Fator2.

Nos fatores subsequentes as cargas fatoriais maiores correspondem aos cortes cárneos e seus respectivos rendimentos. Isso pode ter ocorrido em função dessas características não apresentarem correlações significativas ($P > 0,05$) com a maioria das características relacionadas ao peso corporal e medidas morfométricas (Tabela 4).

O Fator3 apresentou cargas fatoriais maiores para os cortes cárneos PER e LOM e seus respectivos rendimentos (RPER e RLOM). Este fator foi denominado de “Cortes Nobres”. Como descrito por Carvalho *et al.* (2015), a distribuição dos cortes cárneos nas carcaças é de fundamental importância, já que os cortes têm valores econômicos diferentes. Cezar e Sousa (2007) classificaram as carcaças ovinas em função do rendimento muscular, e destacaram o PER e LOM como cortes de 1^o categoria, ou seja, proporciona um alto rendimento muscular, gerando um valor comercial agregado.

O Fator4 apresentou maiores cargas fatoriais para os cortes cárneos PAL, COS, RPAL e RCOS, o qual foi classificado como “Cortes de 2^o Categoria, denominação também adotada por Cezar e Sousa (2007), em função de caracterizarem-se por seu menor rendimento muscular e maior proporção de osso e/ou gordura.

No Fator5, as cargas fatoriais que apresentaram valores maiores foram PG e PT. Essas variáveis são consideradas medidas circulares, tendo sido então denominado “Medidas Morfométricas Circulares”. Essas medidas são utilizadas com o objetivo de avaliação da conformação da carcaça, sendo influenciadas principalmente pela composição corporal que podem variar em função da raça, peso, idade e regime alimentar ao qual é submetido.

O Fator6 apresentou maiores cargas fatoriais para as variáveis PES e RPES. Com base nas características que apresentaram maiores cargas fatoriais no Fator6, podemos nomeá-los esse fator como “Cortes de 3^a Categoria”, como o fez Cezar e Sousa (2007) devido a maior proporção de osso, gordura e, principalmente, tendões ou “nervos”, sendo comercialmente mais baratos.

4. CONCLUSÕES

A análise fatorial mostrou-se adequada para avaliação da carcaça de ovinos da raça Santa Inês.

Dentre as rotações ortogonais e oblíquas, a rotação ortogonal Varimax foi a que apresentou estrutura mais simples para a interpretação dos fatores construídos.

5. REFERÊNCIAS

- BARBOSA, L. *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.
- CAIN, M. K.; ZHANG, Z.; YUAN, K. H. Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. **Behavior Research Methods**, v. 49, n. 5, p. 1716–1735, 2017.
- CARVALHO, Z. G. *et al.* Cortes cárneos e constituintes não-carcaça de ovelhas terminadas em pasto com teores diferentes de suplementação. **Semina: Ciências Agrárias**, v. 36, n. 1, p. 409–420, 2015.
- CEZAR, M. F.; SOUSA, A. H. **Carcaças Ovinas e Caprinas**. 1 ed. João Pessoa: UFCG, 2007.
- CONSTANTIN, C. Principal component analysis - A powerful tool in computing marketing information. **Bulletin of the Transilvania University of Brasov. Series V: Economic Sciences**, v. 7, n. 2, p. 25–30, 2014.
- FINCH, W. H. A Comparison of Factor Rotation Methods for Dichotomous Data. **Journal of Modern Applied Statistical Methods**, v. 10, n. 2, p. 549–570, 2011.

- HAIR, JR. F. *et al.* **Multivariate data analysis**. 6 ed. Reino Unido: Pearson Prentice Hall, 2006.
- HAIR JR, J. *et al.* **Análise Multivariada de Dados**. 6 ed. Porto Alegre: Bookman.
- HOWARD, M. C. A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve? **International Journal of Human-Computer Interaction**, v. 32, n. 1, p. 51–62, 2016.
- KRABBE, P. **The measurement of health and health status: concepts, methods and applications from a multidisciplinary perspective**. Academic Press, 2016.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005.
- OSBORNE, J. W. What is rotating in exploratory factor analysis? **Practical Assessment, Research and Evaluation**, v. 20, n. 2, p. 1–7, 2015.
- ROSA, B. L. *et al.* Correlações entre medidas corporais e características das carcaças de tourinhos Nelore terminados em confinamento. **Boletim de Indústria Animal**, v. 71, n. 4, p. 371–380, 2014.
- SASS, D. A.; SCHMITT, T. A. A comparative investigation of rotation criteria within exploratory factor analysis. **Multivariate Behavioral Research**, v. 45, n. 1, p. 73–103, 2010.
- TABACHNICK, B. G.; FIDELL, L. S. **Using Multivariate Statistics**. 2 ed. Reino Unido, Pearson, 2018.
- TREIBLMAIER, H.; FILZMOSER, P. Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in IS research. **Information and Management**, v. 47, n. 4, p. 197–207, 2010.
- WILLIAMS, M. B.; ONSMAN, A.; BROWN, T. Exploratory factor analysis: A five-step guide for novices. **Journal of Emergency Primary Health Care**, v. 8, n. 3, p. 1–13, 2010.
- YOO, W. *et al.* A Study of Effects of MultiCollinearity in the Multivariable Analysis. **International Journal of Applied Science and Technology**, v. 5, n. 5, p. 9–19, 2014.

CAPÍTULO III

Modelagem multivariada para estimar a composição dos tecidos da carcaça de ovinos da raça Santa Inês

MODELAGEM MULTIVARIADA PARA ESTIMAR A COMPOSIÇÃO DOS TECIDOS DA CARÇA DE OVINOS DA RAÇA SANTA INÊS

RESUMO – O objetivo foi estabelecer um modelo multivariado utilizando duas técnicas estatísticas multivariadas complementares, Análise Fatorial e Regressão Múltipla *Stepwise* para prever a composição dos tecidos por meio das características de carcaça de ovinos da raça Santa Inês. Os dados foram obtidos de 82 ovinos da raça Santa Inês submetidos ao regime de confinamento. As variáveis preditoras foram características de carcaça relacionadas ao peso, rendimento, medidas morfométricas e cortes cárneos. A utilização das variáveis latentes da análise fatorial em modelos de regressão múltipla elimina o problema de multicolinearidade das variáveis explicativas, melhorando assim a acurácia da interpretação dos resultados por propor um melhor ajuste do modelo matemático. Porém, os valores dos coeficientes de determinação (R^2) foram moderados para proporção muscular e gordura total e baixo para a proporção óssea, indicando que variáveis independentes mais adequadas devem ser usadas para melhor prever a proporção dos tecidos em ovinos da raça Santa Inês.

Palavras-chave: Análise multivariada. Dissecção da carcaça. Escores fatoriais. Medidas corporais

MULTIVARIATE MODELING TO ESTIMATE THE COMPOSITION OF CARCASS TISSUES OF SANTA INÊS BREED SHEEP

ABSTRACT – The objective was to establish a multivariate model using two complementary multivariate statistical techniques, Factor Analysis and Stepwise Multiple Regression, to predict tissue composition through carcass characteristics of Santa Inês sheep. The predictor variables were carcass characteristics related to weight, yield, morphometric measurements and meat cuts. The use of latent variables from factor analysis in multiple regression models eliminates the problem of multicollinearity of the explanatory variables, thus improving the accuracy of the interpretation of results by proposing a better adjustment of the mathematical model. However, the values of the coefficients of determination (R^2) were moderate for muscle and total fat proportions and low for bone proportions, indicating that more adequate independent variables should be used to better predict the proportion of tissues in Santa Inês sheep.

Keywords: Multivariate analysis; Carcass dissection; Factor scores; Body measurements

1. INTRODUÇÃO

A avaliação da carcaça é um processo chave para determinar o valor e as características de qualidade dos animais de produção destinados ao abate. Em grande parte, o valor comercial está relacionado ao rendimento e qualidade de carcaça. Como descrito por Ekiz *et al.* (2020), o rendimento e a composição da carcaça (proporções de músculo, gordura e osso) são importantes determinantes da qualidade da carcaça devido à alta variabilidade observada nessas características e seus efeitos evidentes no valor comercial. No entanto, para determinar a composição de forma mais precisa requer a dissecção total ou parcial desses componentes, que é um método caro e demorado.

Nesse sentido, a análise de regressão múltipla é um modelo de predição comumente usada para a interpretação entre uma variável dependente e duas ou mais variáveis independentes, porém, esse método apresenta algumas desvantagens. O desenvolvimento de modelos de regressão múltipla usando variáveis independentes com correlações altas pode apresentar limitações em sua inferência e precisão, e é provável que apresentem sérios efeitos nas estimativas dos coeficientes de regressão e na aplicabilidade geral do modelo estimado (GOMES *et al.*, 2013), devido ao problema da multicolinearidade.

Para evitar esse problema, estudos vêm sendo realizados com o uso dos escores fatoriais ortogonais (variáveis latentes) presente na análise fatorial multivariada. Çelik *et al.* (2018) utilizaram os escores fatoriais para avaliar a influência dos pesos das partes da carcaça sobre o peso total em perus. Daskiram *et al.* (2017), determinaram por intermédio dos escores fatoriais a relação entre produção diária de leite e características do úbere em cabras. Önk *et al.* (2018), Tahtali *et al.* (2019) e Tariq *et al.* (2014) também utilizaram os escores fatoriais para estimar os peso corporal em cordeiros.

A predição da composição corporal e da carcaça de ruminantes foi proposto pela primeira vez por Hankis e Howe (1946), ao qual mostraram que a composição química da secção da 9, 10 e 11^a costelas estava significativamente correlacionada com a composição da carcaça em bovinos de corte. Isso desencadeou estudos subsequentes que mostraram que diferentes cortes e medidas da carcaça podem ser preditoras eficazes na avaliação da composição corporal e da carcaça em geral dos ruminantes (FERNANDES *et al.*, 2008; LAMBE *et al.*, 2009; MARCONDES *et al.*, 2012; RIBEIRO; TEDESCHI, 2012).

Portanto, hipotetizamos que os escores fatoriais ortogonais oriundos da combinação das diferentes características de carcaça podem produzir previsões confiáveis a composição dos tecidos em ovinos da raça Santa Inês. Com base no exposto, objetivou-se estabelecer um modelo multivariado utilizando duas técnicas estatísticas multivariadas complementares, Análise Fatorial e Regressão Múltipla *Stepwise* para prever a composição dos tecidos em ovinos da raça Santa Inês, usando as características de peso, rendimento, medidas morfométricas e cortes cárneos como variáveis independentes.

2. MATERIAL E MÉTODOS

2.1 Experimento e animais

Os experimentos foram realizados no setor de Caprinocultura e Ovinocultura pertencente ao Centro de Ciências Humanas, Sociais e Agrárias da Universidade Federal da Paraíba, localizado no município de Bananeiras, estado da Paraíba, microrregião do Brejo

Paraibano. Foram utilizados 82 ovinos da raça Santa Inês, oriundos de dois experimentos conduzidos para determinar as características de carcaça e qualidade de carne submetidos ao regime de confinamento. Os protocolos de pesquisa dos dois experimentos foram aprovados pelo Comitê de Ética da Universidade Federal da Paraíba.

O experimento 1 teve como objetivo avaliar diferentes níveis de inclusão de palma na dieta e restrição do consumo voluntário de água sobre o desempenho, características de carcaça e qualidade de carne de ovinos da raça Santa Inês. O experimento 2 teve como objetivo avaliar características de carcaça e qualidade da carne em ovinos da raça Santa Inês alimentados com níveis crescentes de resíduo agroindustrial da goiaba (RAG) na dieta. As principais informações dos experimentos estão apresentadas na Tabela 1.

Tabela 1. Principais informações dos experimentos

Características	Experimento 1	Experimento 2
Número de ovinos	42	40
Idade inicial, dias	180	120
Peso inicial, kg	21,6 ± 2,2	21,3 ± 2,2
Peso de abate, kg	28,9 ± 2,6	34,2 ± 3,4
Peso de carcaça fria, kg	14,1 ± 1,3	15,6 ± 1,8

2.2 Procedimentos de abate e características de carcaça

O abate foi procedido em concordância com as normas vigentes do RIISPOA (2000); os animais foram insensibilizados por pistola de dardo cativo, por atordoamento seguido de sangria, por quatro minutos, pela seção da carótida e da jugular. O sangue foi recolhido em recipiente previamente tarado para posterior pesagem.

Após esfolagem e evisceração, foram retiradas a cabeça (secção na articulação atlanto-occipital) e as patas (secção nas articulações metacarpianas e metatarsianas) e registrado o peso da carcaça quente (PCQ). Os componentes internos das cavidades pélvica, abdominal e torácica foram extraídos e tiveram seus pesos registrados. Depois de obtidos o peso de carcaça quente (PCQ), as carcaças foram conduzidas à câmara fria, com temperatura média de 4°C, na qual permaneceram por 24 horas suspensas em ganchos pelo tendão do músculo gastrocnêmico, e posteriormente obtendo o peso de carcaça fria (PCF), segundo metodologia de Cézar e Souza (2007).

O trato gastrointestinal (TGI) foi pesado cheio e vazio para determinação do peso do corpo vazio (PCV, kg) e do rendimento biológico (RB %) = $PCQ/PCV \times 100$. Foram retirados os rins e a gordura perirrenal, que foram subtraídas do PCQ e PCF para calcular os rendimentos de carcaça quente (RCQ, %) = $PCQ/PVA \times 100$, rendimento de carcaça fria (RCF, %) =

$PCF/PVA \times 100$ e o cálculo da perda por resfriamento ($PR, \% = PCQ - PCF/PCQ \times 100$, segundo Cézar e Souza (2007).

Após o período de resfriamento as carcaças foram seccionadas ao meio e as meias-carcaças foram pesadas. Na meia-carcaça esquerda foram mensurados o comprimento interno e externo, comprimento de perna, perímetro do tórax, perímetro da garupa, profundidade do tórax, largura do tórax e largura de garupa, segundo metodologia proposta por Cezar e Sousa (2007). Também foram calculados os índices de compacidade da carcaça (ICC) mediante a fórmula: $ICC (kg/cm) = PCF/\text{comprimento interno de carcaça}$, de acordo metodologia proposta por Cezar e Sousa (2007).

Após as carcaças serem divididas longitudinalmente, as meias carcaças foram seccionadas em cinco regiões anatômicas que compunham os cortes comerciais, metodologia adaptada de Cezar e Sousa (2007), compostas pelo pescoço, paleta, costela, lombo e pernil. Em seguida, o peso individual de cada corte, composto pelos cortes efetuados na meia-carcaça esquerda, foram registrados para cálculo da sua proporção em relação à soma da meia carcaça reconstituída, obtendo-se, assim, o rendimento dos cortes da carcaça.

Os pernis esquerdos de cada animal foram acondicionados em saco de polietileno de alta densidade e congelados a -18°C para avaliação da composição tecidual. Para determinação desta composição, as pernas esquerdas de cada animal foram dissecadas, conforme metodologia descrita por Brown e Williams (1979), previamente descongeladas gradativamente sendo mantidas à temperatura de aproximadamente 4°C durante 24 horas.

Com o auxílio de bisturi, pinça e tesoura foram separados os seguintes grupos tissulares: gordura subcutânea, gordura intermuscular (toda gordura localizada abaixo da fáscia profunda, associada aos músculos), músculo (peso total dos músculos dissecados após remoção completa de toda gordura intermuscular aderida), osso (peso total dos ossos do pernil) e outros tecidos (todos os tecidos não identificados, compostos por tendões, glândulas, nervos e vasos sanguíneos). Por meio da dissecação do pernil foram obtidos os pesos e rendimento dos tecidos dissecados, sendo que a porcentagem dos componentes teciduais foi calculada em relação ao peso reconstituído do pernil, após a dissecação.

2.3 Análises estatísticas

Estatísticas descritivas (média, desvio padrão, variância, valores mínimo e máximo) foram determinadas para todas as variáveis. A análise de correlação de Pearson foi usada para determinar o coeficiente de correlação das composições dos tecidos dissecados (músculos,

ossos e gordura) com as variáveis independentes. As regressões foram desenvolvidas com o PROC REG do SAS® OnDemand for Academics.

A eficácia da análise de regressão múltipla foi determinada usando o coeficiente de determinação (R^2), erro do quadrado médio (MSE), fator de inflação de variância (VIF) como um indicador de multicolinearidade e a estatística C_p de Mallows. O VIF é um indicador de multicolinearidade e indica o quanto um coeficiente de regressão é aumentado devido às correlações entre os preditores do modelo.

O C_p de Mallows (MALLOWS, 2000) é uma medida de qualidade de ajuste, que é frequentemente usada para avaliar o modelo de regressão (MIYASHIRO; TAKANO, 2015). O C_p de Mallows é dado por: $(\frac{SQR}{\sigma^2} + 2p - n)$ (Equação 1), em que SQR é o quadrado médio residual, p é o número de parâmetros no modelo (incluindo o intercepto) e, n é o número de variáveis. O objetivo é encontrar o melhor modelo envolvendo um subconjunto de preditores. Assim, somente aqueles modelos que possuem valores de C_p próximo ao número de parâmetros (incluindo o intercepto) devem ser considerados como um critério desejado para a seleção um subconjunto de preditoras (MALLOWS, 2000).

A análise de regressão múltipla foi usada para estimar as composições dos tecidos dissecados a partir de diferentes características de carcaça. Entretanto, no conjunto de variáveis independentes pode haver variáveis que pouco influenciam as variáveis dependentes; nesse sentido, o procedimento *stepwise* foi usado para selecionar quais variáveis mais influenciam sobre as variáveis dependentes podendo, assim, diminuir o número de variáveis a compor na equação do modelo (ALVES; LOTUFO; LOPES, 2013). De acordo com Senra *et al.* (2007), esse procedimento se baseia na observação de que algumas variáveis contribuem pouco para a eficiência média do modelo; sendo assim, uma vez identificadas, podem ser retiradas do modelo.

A análise de regressão múltipla pelo método *stepwise* foi realizada usando o modelo: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$ (Equação 2), onde Y é a variável dependente ou resposta; α é o intercepto da equação de regressão, β_1 , β_2 e β_n são coeficientes de regressão das variáveis X_1 , X_2 e X_n ; são as variáveis independentes ou variáveis explicativas e e = erro aleatório residual. O critério usado para entrada e permanência de uma variável independente no modelo foi $P > 0,05$.

Na análise de regressão múltipla, um método de estimativa baseado nos escores fatoriais da análise fatorial pode ser usado para eliminar as limitações causadas pelo problema de multicolinearidade entre as variáveis independentes. O principal objetivo da análise fatorial é

permitir compreender e interpretar a suposta relação entre múltiplas variáveis e representar essas múltiplas variáveis em um número mínimos de fatores (variáveis latentes) necessários para a máxima variância representada pelas variáveis originais.

O teste de Bartlett e o teste Kaiser-Meyer-Olkin (KMO) são aplicados para testar a divisibilidade da matriz de correlação em fatores (TAHTALI, 2019). Se a hipótese nula for rejeitada de acordo com os resultados do teste de Bartlett, a análise fatorial pode ser continuada. A obtenção de um índice abaixo de 0,5 com o teste KMO indica que a relação entre os pares de variáveis não pode ser explicada por outras variáveis (ÇELIK *et al.*, 2018), mostrando-se inadequada. Para a interpretação dos fatores, a rotação ortogonal Varimax foi empregada para melhorar a interpretação dos fatores extraídos.

O procedimento PROC FACTOR foi utilizado pelo programa estatístico SAS[®] OnDemand for Academics, para a retenção dos fatores por meio da análise de componentes principais e a escolha do número de fatores pelo critério de Kaiser (1974), que considera os autovalores ≥ 1 como significativos.

O modelo usado na análise foi (Equação 3):

$$X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2$$

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_p$$

O modelo fatorial construído pelos fatores F_1, \dots, F_m , $m \leq p$. Em que: X_1, X_2 e X_p = variáveis em estudo; a = cargas fatoriais; $F = F_1, F_2, \dots, F_m$ são os fatores não correlacionados; $\varepsilon = \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ são variáveis com média 0 e variância 1.

Assim, a análise de regressão múltipla também foi usada para estimar a composição dos tecidos da carcaça a partir dos fatores extraídos, conforme o modelo: $Y = \alpha + \beta_1F_1 + \beta_2F_2 + \dots + \beta_nF_n + e$ (Equação 4), em que Y é a variável dependente/resposta; α é a interceptação da equação de regressão; β_1, β_2 e β_n são coeficientes de regressão dos escores F_1, F_2 e F_n são as variáveis ou fatores explicativos e e = erro aleatório residual.

3. RESULTADOS E DISCUSSÃO

3.1 Estatísticas descritivas variáveis dependentes e independentes

As proporções de tecidos dissecados das carcaças ovinas estão apresentadas na Tabela 2. As proporções de músculo, osso e gordura total foram 68,14, 18,87 e 9,36, respectivamente. Esses valores foram consistentes com os relatados anteriormente para carcaças de ovinos da

raça Santa Inês (CARDOSO *et al.*, 2021; FERNANDES *et al.*, 2021). Proporções semelhantes de tecidos também foram relatadas para ovinos sem padrão racial definido por Lima Júnior *et al.* (2017).

Tabela 2. Estatísticas descritivas variáveis dependentes

Variáveis	$\mu \pm sd$	Variância	Mín.	Máx.
Proporção muscular (%)	68,14 \pm 3,88	15,04	58,87	77,64
Proporção óssea (%)	18,87 \pm 2,26	5,09	15,28	27,81
Proporção de gordura total (%)	9,36 \pm 2,85	8,14	4,10	17,03

$\mu \pm sd$ = média \pm desvio padrão

As estatísticas descritivas para variáveis independentes usadas para prever a composição dos tecidos da carcaça estão apresentadas na Tabela 3. As variáveis apresentadas estão relacionadas as características de peso, rendimento e cortes cárneos. Estudo realizados mostraram que diferentes partes e medidas de carcaças podem ser preditoras precisas na composição corporal e da carcaça em animais ruminantes, considerando-as como variáveis independentes (LAMBE *et al.*, 2009; MARCONDES *et al.*, 2012; RIBEIRO; TEDESCHI, 2012).

Tabela 3. Estatísticas descritivas variáveis independentes

Variáveis	Abreviações	$\mu \pm sd$	Variância	Mín.	Max.
<i>Peso (kg)</i>					
Peso vivo ao abate	PVA	31.58±4.00	15.97	22.70	42.00
Peso de corpo vazio	PCV	25.73±3.33	11.06	18.27	34.70
Peso de carcaça quente	PCQ	15.17±1.80	3.25	11.20	20.20
Peso de carcaça fria	PCF	14.87±1.73	2.98	11.09	20.00
<i>Rendimento (%)</i>					
Rendimento biológico	RB	59.15±3.68	13.51	48.12	73.88
Rendimento de carcaça quente	RCQ	48.18±3.07	9.45	38.64	60.26
Rendimento carcaça fria	RCF	47.27±3.29	10.82	37.08	59.60
Rendimento pernil	RPER	30.08±3.75	14.05	19.37	41.27
Rendimento lombo	RLOM	12.93±2.17	4.70	7.63	20.12
Rendimento costela	RCOS	29.99±3.89	15.16	19.74	38.53
Rendimento pescoço	RPES	6.92±1.19	1.41	4.18	9.98
Rendimento paleta	RPAL	18.81±2.44	5.93	10.90	23.78
<i>Medidas morfométricas (cm)</i>					
Comprimento externo da carcaça	CEC	56.39±3.03	9.15	48.00	65.00
Comprimento interno da carcaça	CIC	60.66±4.59	21.04	51.00	71.00
Comprimento da perna	CP	34.57±2.85	8.12	28.00	39.00
Largura da garupa	LG	17.99±3.47	12.07	12.20	26.00
Largura do tórax	LT	17.21±4.08	16.61	11.20	24.50
Perímetro da garupa	PG	57.09±5.47	29.93	49.00	91.00
Perímetro do pernil	PP	37.31±3.09	9.53	28.00	43.00
Perímetro do tórax	PT	67.96±3.77	14.22	58.00	85.50
Índice de compacidade da carcaça	ICC	0.25±0.02	0.00	0.20	0.29
<i>Cortes (kg)</i>					
Pernil	PER	2.22±0.30	0.09	1.41	3.02
Lombo	LOM	0.96±0.18	0.03	0.64	1.46
Costela	COS	2.22±0.30	0.09	1.51	3.14
Pescoço	PES	0.76±0.26	0.07	0.38	1.27
Paleta	PAL	1.39±0.17	0.03	0.99	1.83

3.2 Coeficientes de correlação entre as características de carcaças e composição dos tecidos

Os coeficientes de correlação entre as proporções dos tecidos da carcaça e características das carcaças são apresentados na Tabela 4. A proporção muscular teve coeficientes de correlação positivos com as características relacionadas as variáveis de rendimento (RB, RCQ, RCF, RCOS, RPES e RPAL), PP e ICC. As correlações mais fortes da proporção muscular foram obtidas com as variáveis de medidas CIC ($r = -0,590$), CP ($r = -0,588$), CP ($r = -0,558$), LG ($r = -0,688$), LT ($r = -0,661$) e os cortes cárneos LOM ($r = -0,570$) e PAL ($r = -0,672$). Ekiz *et al.* (2020), avaliando cabritos da raça caprina Gokceada, não encontraram resultado

significativo ($P>0,05$) entre a variável CIC e proporção muscular. Por outro lado, as correlações da proporção muscular com certas características de carcaça, como PCF, PG, PT, COS, PAL, não foram significativas ($P>0,05$).

Tabela 4. Correlação de Pearson (r) entre a composição dos tecidos e variáveis de carcaça

Variáveis	Abreviações	Músculo (%)	Osso (%)	Gordura (%)
<i>Peso (kg)</i>				
Peso vivo ao abate	PVA	-0.412*	-0.362*	0.501*
Peso de corpo vazio	PCV	-0.428*	-0.364*	0.525*
Peso de carcaça quente	PCQ	-0.255*	-0.361*	0.412*
Peso de carcaça fria	PCF	-0.215 ^{NS}	-0.327*	0.367*
<i>Rendimento (%)</i>				
Rendimento biológico	RB	0.418*	0.069 ^{NS}	-0.331*
Rendimento de carcaça quente	RCQ	0.356*	0.048 ^{NS}	-0.246*
Rendimento carcaça fria	RCF	0.407*	0.113 ^{NS}	-0.319*
Rendimento pernil	RPER	-0.096 ^{NS}	0.094 ^{NS}	0.033 ^{NS}
Rendimento lombo	RLOM	-0.525*	-0.142 ^{NS}	0.494*
Rendimento costela	RCOS	0.222*	0.155 ^{NS}	-0.206 ^{NS}
Rendimento pescoço	RPES	0.371*	-0.091 ^{NS}	-0.235 ^{NS}
Rendimento paleta	RPAL	0.307*	0.093 ^{NS}	-0.282*
<i>Medidas morfológicas (cm)</i>				
Comprimento externo da carcaça	CEC	-0.411*	-0.002 ^{NS}	0.353*
Comprimento interno da carcaça	CIC	-0.590*	-0.186 ^{NS}	0.567*
Comprimento da perna	CP	-0.588*	-0.079 ^{NS}	0.486*
Largura da garupa	LG	-0.688*	-0.313*	0.618*
Largura do tórax	LT	-0.661*	-0.348*	0.605*
Perímetro da garupa	PG	-0.204 ^{NS}	-0.199 ^{NS}	0.205 ^{NS}
Perímetro do pernil	PP	0.490*	0.039 ^{NS}	-0.435*
Perímetro do tórax	PT	-0.180 ^{NS}	-0.271*	0.270*
Índice de compactidade da carcaça	ICC	0.258*	-0.273*	-0.039 ^{NS}
<i>Cortes (kg)</i>				
Pernil	PER	-0.239*	-0.201 ^{NS}	0.324*
Lombo	LOM	-0.570*	-0.324*	0.643*
Costela	COS	0.061 ^{NS}	-0.144 ^{NS}	0.097 ^{NS}
Pescoço	PES	-0.672*	-0.446*	0.746*
Paleta	PAL	0.161 ^{NS}	-0.212 ^{NS}	0.023 ^{NS}

^{NS} $P>0,05$; * $P<0,05$.

A proporção óssea apresentou coeficientes de correlações negativas significativas ($P<0,05$) com as características associadas ao peso (PG, LT, PT, ICC LOM e PES); as demais variáveis não apresentaram correlações significativas ($P>0,05$) (Tabela 4). Esses resultados estão de acordo com Díaz *et al.* (2004) e Ekiz *et al.* (2020), que relataram que a proporção óssea

teve correlações negativas e significativas com características associadas ao peso, medidas e índices de carcaça.

Com exceção para PG, ICC, COS, RPER, RLOM, RCOS e RPES, as demais variáveis apresentaram correlações significativas com a proporção de gordura ($P < 0,05$) (Tabela 4). Dentre essas características, as que apresentaram maiores correlações foram LG, PG, LOM e PAL, assim como Santos *et al.* (2017), que relataram correlações significativas e positivas da proporção de gordura com peso de carcaça fria, medidas de carcaças e cortes cárneos na carcaça de cabritos lactantes.

Independentemente de as correlações serem positivas ou negativas, podemos observar que a maioria foi significativa ($P < 0,05$), indicando que essas variáveis podem ser utilizadas como indicadores para a proporção dos tecidos.

3.3 Predição da composição dos tecidos da carcaça usando análise de regressão múltipla *stepwise* das variáveis originais

As equações de regressão para prever a composição dos tecidos por meio da análise de regressão múltipla *stepwise* das variáveis originais estão apresentadas na Tabela 5. De acordo com a análise, as variáveis independentes LG, ICC, LOM e PAL foram as melhores preditoras da proporção muscular, explicando 63,6% da variação da proporção muscular com base no R^2 .

No entanto, os resultados baseados nos valores de C_p indicam que os modelos propostos apresentaram falta de ajuste. Modelos aceitáveis dentro de um subconjunto de variáveis, os valores de C_p precisam estar próximo do número de preditoras mais a constante, isso indica que o modelo é relativamente não-viciado em estimar os coeficientes verdadeiros de regressão e prever respostas futuras (KAZEMI *et al.*, 2013). Tais resultados podem estar diretamente relacionados à presença de multicolinearidade entre as variáveis independentes. As variáveis usadas nos modelos apresentaram de forma moderadamente correlacionadas ($1 < \text{VIF} < 5$); nesse sentido, o uso das variáveis de carcaças deve ser usado com cautela, pois a multicolinearidade está associada a estimativas instáveis dos coeficientes de regressão.

Em estudos com essas mesmas perspectivas, Díaz *et al.* (2004) determinaram que a melhor equação de predição para a proporção muscular da carcaça de cordeiros lactantes da raça Manchego incluía o peso da gordura do canal do botão do rim, índice de gordura, CIC e peso do osso do canhão dianteiro com variáveis independentes. Em cordeiros da raça Churra Tensina, a equação de regressão incluiu a largura da carcaça e o peso da gordura do canal do botão do rim como variáveis independentes para melhor prever a proporção do músculo (CARRASCO *et al.*, 2009). Ekiz *et al.* (2020) relataram a porcentagem da gordura do canal do

botão do rim como melhor preditor da proporção muscular. As equações relatadas por Díaz *et al.* (2004) e Carrasco *et al.* (2009) obtiveram precisão semelhante em comparação ao estudo atual (valores de $R^2 = 0,63$ e $0,58$, respectivamente) e precisão maior comparado ao estudo de Ekiz *et al.* (2020), avaliando cabritos da raça Gokceada, que encontraram valor de $R^2 = 0,21$. De acordo Ekiz *et al.* (2020), as diferenças entre os estudos em termos de precisão podem ser devido as diferenças de espécies, bem como diferenças no número de variáveis independentes alocadas no modelo.

1 **Tabela 5.** Equações de predição para a composição dos tecidos de acordo a análise de regressão múltipla *stepwise* das variáveis originais.

Variável dependente	Número	Equação	C_p	R^2	MSE	Valor P
Proporção muscular (%)	1	$Y = 91,961 - 0,768 \times \mathbf{LG}$	27,290	0,474	2,831	<.0001
	2	$Y = 68,314 - 0,798 \times \mathbf{LG} + 57,885 \times \mathbf{ICC}$	9,595	0,574	2,563	<.0001
	3	$Y = 69,817 - 0,653 \times \mathbf{LG} + 57,293 \times \mathbf{ICC} - 4,227 \times \mathbf{LOM}$	7,392	0,596	2,514	<.0001
	4	$Y = 68,151 - 0,493 \times \mathbf{LG} + 39,931 \times \mathbf{ICC} - 9,014 \times \mathbf{LOM} + 6,582 \times \mathbf{PAL}$	1,438	0,636	2,401	<.0001
Proporção óssea (%)	1	$Y = 21,853 - 3,940 \times \mathbf{PES}$	16,141	0,199	2,032	<.0001
	2	$Y = 29,708 - 31,432 \times \mathbf{ICC} - 4,076 \times \mathbf{PES}$	8,044	0,287	1,930	<.0001
	3	$Y = 21,476 + 0,264 \times \mathbf{CP} - 29,227 \times \mathbf{ICC} - 5,984 \times \mathbf{PES}$	2,676	0,351	1,853	<.0001
	4	$Y = 25,638 + 0,331 \times \mathbf{CP} - 0,117 \times \mathbf{PT} - 23,331 \times \mathbf{ICC} - 5,927 \times \mathbf{PES}$	1,411	0,379	1,824	<.0001
	5	$Y = 25,682 + 0,374 \times \mathbf{CP} + 0,087 \times \mathbf{PG} - 0,189 \times \mathbf{PT} - 28,207 \times \mathbf{ICC} - 6,420 \times \mathbf{PES}$	0,639	0,403	1,801	<.0001
	6	$Y = 21,476 + 0,110 \times \mathbf{RCF} + 0,404 \times \mathbf{CP} + 0,093 \times \mathbf{PG} - 0,208 \times \mathbf{PT} - 33,589 \times \mathbf{ICC} - 6,094 \times \mathbf{PES}$	0,512	0,422	1,785	<.0001
Proporção gordura total (%)	1	$Y = 3,052 + 8,331 \times \mathbf{PES}$	2,420	0,556	3,660	<.0001
	2	$Y = 5,718 + 8,137 \times \mathbf{PES} - 0,364 \times \mathbf{RPES}$	0,429	0,579	3,518	<.0001

2 C_p de Mallows (C_p); Coeficiente de determinação (R^2); Erro do quadrado médio (MSE); Largura da garupa (LG); Índice de compactidade da carcaça (ICC);
3 Lombo (LOM); Paleta (PAL); Pescoço (PES); Comprimento da perna (CP); Perímetro do tórax (PT); Perímetro da garupa (PG); Rendimento da carcaça fria
4 (RCF); Rendimento do pescoço (RPES).

5

O CP, PG, PT, ICC e PES foram incluídos na equação de predição da proporção óssea (Tabela 5). Essas cinco variáveis explicaram apenas 40,3% da variação da proporção óssea, o que indica que a previsão desta proporção usando este modelo foi baixa. Nesse sentido, seria de esperar que dentre os modelos propostos este poderia ser o melhor para prever a predição da proporção óssea. Porém, com base nos valores de C_p , o modelo melhor ajustado foi para a equação que considerou apenas as variáveis CP, ICC e PES como preditoras. Embora existam divergências na literatura em decidir qual variável individual é mais adequada para ser utilizada na predição de carcaça de animais, a precisão da predição em relação ao R^2 tem sido melhorada principalmente quando mais de uma variável é considerada no modelo. No atual estudo ficou demonstrado que nem sempre a adição de mais variáveis preditoras e melhora do R^2 produz um aumento na precisão das estimativas obtidas.

Em um estudo anterior conduzido para prever a proporção óssea de cordeiros da raça Churra Tensina, Carrasco *et al.* (2009) descobriram que a melhor equação obtida incluía peso da gordura do canal do botão do rim e pontuação de conformação ($R^2 = 0,51$). A equação relatada por Díaz *et al.* (2004) incluiu o escore de gordura e a proporção de gordura omental para prever a proporção óssea de cordeiros lactantes da raça Manchego ($R^2 = 0,76$). Por outro lado, Ekiz *et al.* (2020) incluíram a compacidade dos membros posteriores e o peso da cauda na equação de predição de proporção óssea ($R^2 = 0,62$).

De acordo com os resultados da análise de regressão múltipla *stepwise* das variáveis originais, duas variáveis (PES e RPES) foram determinadas para prever a proporção da gordura total. Essa equação de predição explicou 57,9% da proporção da gordura total. Díaz *et al.* (2004), Carrasco *et al.* (2009) e Ekiz *et al.* (2020) predisseram a proporção de gordura total com acurácias superiores ($R^2 = 0,84, 0,73$ e $0,68$, respectivamente). No estudo desenvolvido por Díaz *et al.* (2004), a equação foi obtida com o modelo incluindo variáveis de gordura (espessura de gordura, escore de gordura e porcentagem da gordura do canal do botão do rim) para a predição da proporção de gordura. No estudo de Carrasco *et al.* (2009), o modelo da equação incluiu peso da gordura do canal do botão do rim, largura da carcaça e comprimento interno da carcaça como variáveis independentes.

No estudo atual, a proporção óssea foi prevista como menos precisa (menor R^2) do que as proporções de músculo e gordura (Tabela 5). Ao contrário de Ekiz *et al.* (2020), que descobriram que a quantidade de variação explicada foi menor na equação de predição para a proporção muscular do que osso e gordura. Quanto às equações de regressão para prever as composições dos tecidos, vale ressaltar que as medidas morfométricas, exceto, LG, CP e PT, as demais não foram admitidas nos modelos. Em estudo realizado com cabritos da raça

Gokceada, Ekiz *et al.* (2020) também observou a falta de relação entre medidas de dimensões da carcaça e a composição dos tecidos. De acordo Cadavez (2009), as medidas de carcaça refletem a dimensão do esqueleto em vez da composição dos tecidos da carcaça.

Com base nas correlações observadas entre composição dos tecidos com as características de carcaça (Tabela 4), seria de esperar que outras medidas da carcaça fossem incluídas nas equações de predição. Nesse caso, poderíamos sugerir o problema de multicolinearidade, uma vez que no estudo atual foram observados problemas de multicolinearidade moderada ($1 < VIF < 5$).

De acordo com os presentes resultados, as equações de predição obtidas pela regressão das variáveis originais parecem ser insuficientes para predizer com precisão a composição dos tecidos, porque a quantidade de variação explicada foi baixa para proporção óssea e moderada para proporção muscular e de gordura. Por outro lado, as medidas lineares e circulares da carcaça consideradas no estudo foram excluídas dos modelos durante os procedimentos de seleção da regressão de *stepwise*, o que indica que essas variáveis foram fracas preditoras para as proporções de músculo, osso e gordura total em ovinos da raça Santa Inês.

3.4 Predição da composição dos tecidos da carcaça usando regressão de *stepwise* das variáveis latentes

Os resultados do teste de esfericidade de Barlett e Kaiser-Meyer-Olkin (KMO) estão apresentados na Tabela 6, que são pré-requisitos essenciais para a análise fatorial. Em outras palavras, a obtenção de resultados significativos de ambos os testes mostra que os dados são adequados para a análise fatorial. Considerando os valores obtidos pelo teste de Bartlett ($P < 0,05$) e KMO (0,702) para testar a divisibilidade da matriz de correlação dos fatores, os dados foram considerados adequados para a análise fatorial.

Tabela 6. Adequação da amostragem pelo teste de esfericidade de Barlett e o teste de KMO (Kaiser-Meyer-Olkin).

	DF	325
Teste de esfericidade de Bartlett	Chi-Square	4761,91
	Pr > Chi Sq	<.0001
Teste KMO (Kaiser-Meyer-Olkin)	---	0,702

Os valores das cargas fatoriais obtidas pela rotação Varimax, comunalidades, autovalores e variância explicada estão descritos na Tabela 7. Podemos observar que a retenção de fatores por meio da análise de componentes principais e a escolha do número de fatores pelo

critério de Kaiser foi possível extrair seis fatores (autovalores > 1). Esses seis fatores foram capazes de explicar cerca de 85,40% da variância original das variáveis. Esse valor demonstrou-se bastante expressivo, representando pouca perda de informações, indicando que a análise resumiu a maior parte da informação em um número mínimo de fatores necessários para explicar o máximo de variância representada pelas variáveis originais.

Tabela 7. Resultados da análise fatorial aplicada às variáveis independentes.

Variáveis	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	C
PVA	0,762	0,122	-0,404	0,349	0,315	-0,035	0,980
PCV	0,773	0,131	-0,353	0,342	0,348	-0,032	0,978
PCQ	0,783	0,078	0,074	0,466	0,368	-0,017	0,978
PCF	0,779	0,011	0,112	0,469	0,376	-0,052	0,984
RB	-0,102	-0,135	0,904	0,153	-0,033	0,033	0,871
RCQ	-0,046	-0,112	0,956	0,159	0,056	0,035	0,959
RCF	-0,085	-0,215	0,939	0,130	0,045	-0,022	0,955
CEC	0,712	0,319	-0,051	0,002	0,089	-0,185	0,653
CIC	0,600	0,550	-0,184	0,080	0,361	-0,049	0,835
CPE	0,617	0,598	-0,171	-0,010	0,087	-0,051	0,778
LG	0,523	0,473	-0,355	0,121	0,450	0,074	0,846
LT	0,454	0,409	-0,396	0,094	0,526	0,092	0,824
PG	0,179	-0,071	0,015	0,167	0,859	0,063	0,807
PP	-0,123	-0,774	0,091	0,116	0,171	-0,061	0,670
PT	0,329	0,137	0,162	-0,042	0,699	-0,133	0,661
ICC	0,419	-0,546	0,202	0,573	0,230	-0,047	0,898
PER	0,302	0,220	0,113	0,431	0,156	0,770	0,956
LOM	0,235	0,697	-0,034	0,462	0,248	0,340	0,932
COS	0,076	0,212	0,389	0,760	-0,062	0,069	0,789
PES	0,309	0,706	-0,268	0,244	0,347	0,089	0,855
PAL	-0,163	0,028	0,109	0,863	0,131	0,178	0,833
RPER	-0,357	0,229	-0,005	0,000	-0,153	0,872	0,963
RLOM	-0,251	0,789	-0,147	0,188	0,040	0,413	0,916
RCOS	-0,570	0,210	0,280	0,342	-0,387	0,091	0,723
RPES	-0,802	-0,015	-0,005	0,199	0,012	-0,194	0,720
RPAL	-0,791	0,016	-0,010	0,382	-0,186	0,182	0,840
Autovalores	9,802	4,437	4,105	1,702	1,119	1,039	Total
Variância (%)	37,70	17,06	15,79	6,54	4,30	4,00	85,40

Comunalidade (C); Peso de corpo vazio (PCV); Peso vivo ao abate (PVA); Peso de carcaça quente (PCQ); Peso de carcaça fria (PCF); Rendimento biológico (RB); Rendimento de carcaça quente (RCQ); Rendimento de carcaça fria (RCF); Comprimento externo da carcaça (CEC); Comprimento interno da carcaça (CIC); Comprimento da perna (CP); Largura da garupa (LG); Largura do tórax (LT); Perímetro da garupa (PG); Perímetro do pernil (PP); Perímetro do tórax (PT); Índice de compacidade da carcaça (ICC); Pernil (PER); Lombo (LOM); Costela (COS); PESCOÇO (PES); Paleta (PAL); Rendimento do

pernil (RPER); Rendimento do lombo (RLOM); Rendimento da costela (RCOS); Rendimento do pescoço (RPES); Rendimento da paleta (RPAL).

A comunalidade apresentada na Tabela 7 é a proporção de variabilidade de cada variável que é explicada pelos fatores. Quanto mais próximo de 1, melhor a variável explica a variação nos fatores. Hair Jr. *et al.* (2009) descrevem que pelo menos metade da variância de cada variável deve ser levada em conta, usando essa diretriz, todas as variáveis com comunalidade menores que 0,50 apresentam explicação insuficiente. Na Tabela 6, todas as variáveis apresentaram comunalidade maiores que 0,50, indicando que a proporção da variabilidade é adequada para explicar os fatores.

A Tabela 8 ilustra os resultados da análise da regressão múltipla *stepwise* com as novas variáveis independentes (latentes) da análise fatorial. De acordo com a análise, todas as variáveis latentes selecionadas pela análise de componentes principais (Factor1, 2, 3, 4, 5 e 6) foram consideradas como melhores preditoras da proporção muscular, explicando 62,5% da variação da proporção muscular. Porém, a variação da proporção muscular explicada por essas variáveis latentes é inferior ao resultado encontrado na análise de regressão das variáveis originais (Tabela 5), sendo que o melhor modelo (LG, ICC, LOM e PAL) explicou 63,7% da variação para a proporção muscular. No entanto, devemos considerar as variáveis latentes como a equação de melhor ajuste do modelo, em função do valor de C_p apresentar o mesmo número de variáveis preditoras mais a constante (7,000). Tais resultados podem estar diretamente relacionados ao fato da ausência de multicolinearidade entre as variáveis preditoras ($VIF = 1$). De acordo Tariq *et al.* (2012), um dos métodos mais eficazes para resolver o problema da multicolinearidade na análise de regressão múltipla é usar as cargas fatoriais da análise fatorial, pelas variáveis latentes.

Tabela 8. Equações de predição para a composição dos tecidos de acordo a análise de regressão múltipla *stepwise* das variáveis latentes.

Variável dependente	Número	Equação	C(p)	R ²	MSE	Valor P
Proporção muscular, %	1	$Y = 68,142 - 2,328 \times \mathbf{Factor3}$	48,630	0,360	9.745	<.0001
	2	$Y = 68,142 - 1,200 \times \mathbf{Factor1} - 2,328 \times \mathbf{Factor3}$	31,978	0,456	8.394	<.0001
	3	$Y = 68,142 - 1,200 \times \mathbf{Factor1} + 1,138 \times \mathbf{Factor2} - 2,328 \times \mathbf{Factor3}$	17,213	0,542	7.159	<.0001
	4	$Y = 68,142 - 1,200 \times \mathbf{Factor1} + 1,138 \times \mathbf{Factor2} - 2,328 \times \mathbf{Factor3} - 0,795 \times \mathbf{Factor5}$	11,027	0,584	6.589	<.0001
	5	$Y = 68,142 - 1,200 \times \mathbf{Factor1} + 1,138 \times \mathbf{Factor2} - 2,328 \times \mathbf{Factor3} - 0,795 \times \mathbf{Factor5} - 0,581 \times \mathbf{Factor6}$	8,651	0,607	6.317	<.0001
	6	$Y = 68,142 - 1,200 \times \mathbf{Factor1} + 1,138 \times \mathbf{Factor2} - 2,328 \times \mathbf{Factor3} + 0,531 \times \mathbf{Factor4} - 0,795 \times \mathbf{Factor5} - 0,581 \times \mathbf{Factor6}$	7,000	0,625	6,099	<.0001
Proporção óssea, %	1	$Y = 18,870 - 0,746 \times \mathbf{Factor5}$	8,629	0,109	4,590	0,0027
	2	$Y = 18,870 - 0,673 \times \mathbf{Factor4} - 0,746 \times \mathbf{Factor5}$	2,181	0,198	4,186	0,0002
Proporção gordura total, %	1	$Y = 9,359 + 1,518 \times \mathbf{Factor3}$	32,048	0,283	5,907	<.0001
	2	$Y = 9,359 + 0,880 \times \mathbf{Factor1} + 1,518 \times \mathbf{Factor3}$	19,703	0,378	5,189	<.0001
	3	$Y = 9,359 + 0,880 \times \mathbf{Factor1} + 1,518 \times \mathbf{Factor3} + 0,709 \times \mathbf{Factor5}$	12,392	0,440	4,735	<.0001
	4	$Y = 9,359 + 0,880 \times \mathbf{Factor1} - 0,677 \times \mathbf{Factor2} + 1,518 \times \mathbf{Factor3} + 0,709 \times \mathbf{Factor5}$	5,899	0,496	4.315	<.0001

C_p de Mallows (*C_p*); Coeficiente de determinação (R²); Erro do quadrado médio (MSE); Largura da garupa (LG); Índice de compactidade da carcaça (ICC); Lombo (LOM); Paleta (PAL); Pescoço (PES); Comprimento da perna (CP); Perímetro do tórax (PT); Perímetro da garupa (PG); Rendimento da carcaça fria (RCF); Rendimento do pescoço (RPES).

As variáveis latentes Factor4 e Factor5 foram incluídas na equação de predição da proporção óssea (Tabela 8). No entanto, essas duas variáveis latentes explicaram somente 19,8%, valor expressamente inferior ao encontrado na Tabela 5 (35,1%) com as variáveis CP, ICC e PES, que foram selecionadas pelo procedimento de *stepwise* com o melhor modelo desenvolvido para a proporção óssea. Com base nas cargas fatoriais apresentadas na Tabela 7 para as variáveis latentes correspondente ao Factor 1 e 3, as variáveis que mais contribuíram foram aquelas relacionadas ao peso e medidas corporais (CEC, CIC, CP e LG) e as características de rendimento (exceto RPER e RLOM). O ICC, que foi considerado no modelo anterior (Tabela 5) não foi considerado no modelo atual uma vez que não apresentou carga fatorial expressiva no Factor1 e 3.

De acordo com os resultados apresentados na Tabela 8, quatro variáveis latentes (Factor1, 2, 3 e 5) foram determinadas para predizer a proporção da gordura total, explicando 49,6% da proporção da proporção da gordura total. Assim como observado para a proporção muscular e óssea, a proporção da gordura total também apresentou explicação inferior comparado a equação anteriormente proposta na Tabela 5 que apresentou explicação de 57,9%.

Estudos anteriores foram realizados utilizando os escores fatoriais da análise fatorial (variáveis latentes) nas equações da regressão múltipla (ÇELIK *et al.*, 2018; KHAN *et al.*, 2014; ÖNK *et al.*, 2018; TAHTALI, 2019; TARIQ *et al.*, 2012). Os coeficientes de determinação encontrados variavam de 0,754 a 0,966, mostrando-se adequados para serem usados como variáveis independentes na análise de regressão. No estudo atual, os coeficientes de determinação apresentaram valores de baixo a moderado (R^2 variando de 0,198 a 0,607) para a composição dos tecidos. Com esses resultados podemos concluir que os modelos preditivos desenvolvidos não são eficazes para prever a composição dos tecidos da carcaça.

Todavia, para fins de comparação podemos concluir que os modelos desenvolvidos das variáveis latentes promovem um melhor ajuste nas equações com base nos valores de C_p . Apesar de apresentarem menores valores de R^2 , ficou clara a presença de multicolinearidade moderada entre as variáveis originais ($1 < VIF < 5$), indicando um aumento da variância dos coeficientes de regressão devido às correlações entre as preditoras do modelo.

4. CONCLUSÃO

A utilização das variáveis latentes da análise fatorial em modelos de regressão múltipla elimina o problema de multicolinearidade das variáveis explicativas, melhorando, assim, a acurácia da interpretação dos resultados por propor um melhor ajuste do modelo matemático.

Porém, os valores de R^2 foram moderados para proporção muscular e gordura total e baixo para a proporção óssea, indicando que variáveis independentes mais adequadas devem ser usadas para melhor prever a proporção dos tecidos em ovinos da raça Santa Inês.

5. REFERÊNCIAS

- ALVES, M. F.; LOTUFO, A. D. P.; LOPES, M. L. M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. **Proceeding Series of the Brazilian Society of Applied and Computational Mathematics**, v. 1, n. 1, p. 1–6, 2013.
- BROWN, A.; WILLIAMS, D. Sheep carcass evaluation-measurement of composition using a standardised butchery method. **Memorandum-Meat Research Institute**, 1979.
- CADAVEZ, V. A. P. Prediction of lean meat proportion of lamb carcasses. **Archiva Zootechnica**, v. 12, n. 4, p. 46–58, 2009.
- CARDOSO, D. B. *et al.* Growth performance, carcass traits and meat quality of lambs fed with increasing levels of spineless cactus. **Animal Feed Science and Technology**, v. 272, 2021.
- CARRASCO, S. *et al.* Carcass tissue composition in light lambs: Influence of feeding system and prediction equations. **Livestock Science**, v. 126, n. 1–3, p. 112–121, 2009.
- ÇELİK, Ş. *et al.* Analysis of variables affecting carcass weight of white Turkeys by regression analysis based on factor analysis scores and ridge regression. **Revista Brasileira de Ciencia Avicola**, v. 20, n. 2, p. 273–280, 2018.
- CEZAR, M. F.; SOUSA, A. H. **Carcaças ovinas e caprinas**. ed. João Pessoa: UFCG, 2007.
- DASKIRAN, I.; KESKIN, S.; BINGOL, M. Usability of the factor analysis scores in multiple linear regression analyses for the prediction of daily milk yield in Norduz Goats. **Journal of Agricultural Science and Technology**, v. 19, p. 1507–1515, 2017.
- DÍAZ, M. T. *et al.* Prediction of suckling lamb carcass composition from objective and subjective carcass measurements. **Meat Science**, v. 66, n. 4, p. 895–902, 2004.
- EKIZ, B. *et al.* Comparison of the decision tree, artificial neural network and multiple regression methods for prediction of carcass tissues composition of goat kids. **Meat Science**, v. 161, p. 108011, 2020.
- EYDURAN, E. *et al.* Carcass weight estimation from some morphological traits of Capoeira capoeira (Güldenstädt, 1772) using factor scores in multiple regression analysis. **Pakistan Journal of Statistics**, v. 28, n. 1, p. 159–165, 2012.
- FERNANDES, J. *et al.* Carcass and meat quality in lambs receiving natural tannins from *Mimosa tenuiflora* hay. **Small Ruminant Research**, v. 198, 2021.
- FERNANDES, M. H. M. R. *et al.* Predicting the chemical composition of the body and the carcass of 3/4Boer × 1/4Saanen kids using body components. **Small Ruminant Research**, v. 75, n. 1, p. 90–98, 2008.
- GOMES, H. F. B. *et al.* Common factors method to predict the carcass composition tissue in kid goats. **Revista Brasileira de Zootecnia**, v. 42, n. 3, p. 193–203, 2013.
- HAIR JR, J. *et al.* **Análise Multivariada de Dados**. 6.ed. Porto Alegre: Bookman, 2009.
- HANKINS, O. G.; HOWE, P. E. **Estimation of the composition of beef carcasses and cuts**. 1.ed. Los Angeles, 1946.

- KAISER, H. F. An index of factorial simplicity. **Psychometrika**, v. 39, n. 1, p. 31–36, 1974.
- KAZEMI, A. *et al.* Optimal power quality monitor placement using genetic algorithm and Mallow's Cp. **International Journal of Electrical Power and Energy Systems**, v. 53, n. 1, p. 564–575, 2013.
- KHAN, M. A. *et al.* Estimating body weight from several body measurements in Harnai sheep without multicollinearity problem. **Journal of Animal and Plant Sciences**, v. 24, n. 1, p. 120–126, 2014.
- LAMBE, N. R. *et al.* Prediction of lamb carcass composition and meat quality using combinations of post-mortem measurements. **Meat Science**, v. 81, n. 4, p. 711–719, 2009.
- LIMA JÚNIOR, D. M. *et al.* Leg tissue composition and physico-chemical parameters of sheep meat fed annatto coproduct. **Semina: Ciências Agrárias**, v. 38, n. 5, p. 3377–3382, 2017.
- MARCONDES, M. I. *et al.* Prediction of physical and chemical body compositions of purebred and crossbred Nellore cattle using the composition of a rib section. **Journal of Animal Science**, v. 90, n. 4, p. 1280–1290, 2012.
- MIYASHIRO, R.; TAKANO, Y. Subset selection by Mallows' Cp: A mixed integer programming approach. **Expert Systems with Applications**, v. 42, n. 1, p. 325–331, 2015.
- ÖNK, K.; SARI, M.; GÜRCAN, İ. S. Estimation of live weights at the beginning and the end of grazing season in Tuj lambs via scores of factor analysis. **Ankara Üniversitesi Veteriner Fakültesi Dergisi**, v. 65, n. 3, p. 261–266, 2018.
- RIBEIRO, F. R. B.; TEDESCHI, L. O. Using real-time ultrasound and carcass measurements to estimate total internal fat in beef cattle over different breed types and managements. **Journal of Animal Science**, v. 90, n. 9, p. 3259–3265, 2012.
- RIISPOA. **Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal**, 2000. (Nota técnica).
- SANTOS, V. A. C. *et al.* Estimation of carcass composition of goat kids from joint dissection and conformation measurements. **Italian Journal of Animal Science**, v. 16, n. 4, p. 659–665, 2017.
- SENRA, L. F. A. DE C. *et al.* Estudo sobre métodos de seleção de variáveis em DEA. **Pesquisa Operacional**, v. 27, n. 2, p. 191–207, 2007.
- TAHTALI, Y. Use of factor scores in multiple regression analysis for estimation of body weight by certain body measurements in Romanov Lambs. **PeerJ**, v. 2019, n. 8, p. 1–11, 2019.
- TARIQ, M. M. *et al.* Prediction of body weight from testicular and morphological characteristics in indigenous mengali sheep of Pakistan using factor analysis scores in multiple linear regression analysis. **International Journal of Agriculture and Biology**, v. 14, n. 4, p. 590–594, 2012.

CAPÍTULO IV

Diferenciação de ovinos de acordo com escores de conformação e acabamento de carcaça: um estudo usando análise discriminante multivariada

DIFERENCIAÇÃO DE OVINOS DE ACORDO COM ESCORES DE CONFORMAÇÃO E ACABAMENTO DE CARÇAÇA: UM ESTUDO USANDO ANÁLISE DISCRIMINANTE MULTIVARIADA

RESUMO – A análise discriminante é uma técnica multivariada utilizada para classificar e discriminar grupos ou objetos. Esta pesquisa utilizou a análise discriminante multivariada para distinguir grupos de ovinos em função do escore de conformação e acabamento de carcaça, e identificar as variáveis que mais contribuem na diferenciação. O banco de dados foi composto por informações de ovinos não castrados da raça Santa Inês, submetidos ao regime de confinamento. Os dados foram submetidos ao teste de Mardia, para verificar a normalidade multivariada, seguido do teste não-paramétrico k-vizinho mais próximo (KNN). Um subconjunto particular de variáveis foi selecionado pelo procedimento *stepwise* e a distância de Mahalanobis (D^2) utilizada para a separação dos grupos. Os animais avaliados foram classificados 100% em seus grupos de origem. As variáveis com maior poder discriminatório para os escores de conformação de carcaça foram peso de carcaça fria (PCF), comprimento externo da carcaça (CEC) e pescoço (PES) e para acabamento de carcaça foram peso vivo ao abate (PVA), CEC e perímetro torácico (PT). A análise discriminante multivariada mostrou-se eficiente para alocar os animais em seus grupos de origem (escores de carcaça).

Palavras-chave: Análise multivariada. Classificação de carcaça. Seleção *stepwise*.

DIFFERENTIATION OF SHEEP ACCORDING TO CARCASS CONFORMATION AND FINISHING SCORE: A STUDY USING MULTIVARIATE DISCRIMINANT ANALYSIS

ABSTRACT – The carcass classification consists of clustering animals in groups with similar carcass traits. The carcass conformation describes the carcass's desirable shape (convex or concave), representing the proportion of muscle and fat in bone. When groups are defined a priori, as in conformation scores and finishing score, the interest is to identify the contribution of each variable used to separate the groups. Therefore, multivariate discriminant analysis was used to discriminate Santa Inês animals (locally adapted breed from Northeastern Brazil) as a function of carcass conformation scores and finishing score (score 1 = poor, score 2 = reasonable, score 3 = good, score 4 = very good, and 5 = excellent) and identify the variables that most contribute to differentiation. The database was composed of 122 pieces of information from non-castrated Santa Inês sheep, submitted to the confinement regime, of which 24 variables related to the carcass of the animals were recorded. First, data were submitted to the Mardia's test to verify multivariate normality, followed by the non-parametric k-nearest neighbor test. Finally, the stepwise procedure selected a particular subset of variables, and the Mahalanobis distance was used to separate the groups. The evaluated animals were classified 100% in their groups of origin. The variables with the highest discriminatory power for carcass conformation scores were cold carcass weight (CCW), external carcass length (EXL) and neck (NEC) and for carcass finishing were body weight at slaughter (BWS), EXL and chest perimeter (CP). Multivariate discriminant analysis efficiently allocates the animals to their original groups (carcass scores).

Keywords: Multivariate analysis. Carcass classification. *Stepwise* selection.

1. INTRODUÇÃO

Na produção animal, o processo de classificação de carcaças consiste em agrupar em classes as carcaças com características semelhantes como peso, acabamento e conformação da carcaça. Esses dois últimos são ferramentas potenciais para classificar e prever a qualidade da

carcaça. Segundo Jones *et al.* (2021), a conformação indica o formato desejável da carcaça quanto ao perfil desejado (convexo ou côncavo), ao qual representa a proporção de músculo e gordura em relação ao osso. Já o acabamento quantifica a quantidade de gordura subcutânea na carcaça visível ao avaliador e é utilizado para selecionar o mercado de destino, compatível com a preferência do consumidor.

Quando os grupos são definidos *a priori*, como o caso dos escores de conformação e acabamento de carcaça (escore 1 = ruim, escore 2 = razoável, escore 3 = boa, escore 4 = muito boa e 5 = excelente), o interesse é identificar a contribuição de cada variável utilizada para a separação dos grupos. Nesse sentido, a análise discriminante multivariada é empregada na classificação e diferenciação de objetos ou grupos e na seleção de variáveis com maior poder discriminatório (ALKARKHI; ALQARAGHULI, 2020). Em outras palavras, a análise discriminante multivariada é usada para selecionar variáveis contínuas que discriminam entre dois ou mais grupos e determinar combinações lineares das variáveis (JEON *et al.*, 2013), ou seja, variáveis que fornecem a discriminação máxima entre grupos.

Levando em consideração que as pontuações dos escores de conformação e acabamento de carcaça na sua grande proporção são baseadas na avaliação visual da musculatura e da gordura subcutânea, pode ocorrer uma maior incidência de observações classificadas incorretamente, tornando útil a avaliação por meio da análise discriminante multivariada.

Dado o exposto, objetivou-se pela análise discriminante distinguir grupos de conformação e acabamento de carcaça de ovinos, usando um conjunto particular de variáveis que melhor caracterizam ou separam os diferentes grupos em análise.

2. MATERIAL E MÉTODOS

O banco de dados continha informações de carcaças de ovinos não castrados da raça Santa Inês, submetidos ao regime de confinamento. O banco de dados é oriundo de três experimentos conduzidos no setor de Caprinocultura e Ovinocultura pertencente ao Centro de Ciências Humanas, Sociais e Agrárias da Universidade Federal da Paraíba, localizado no município de Bananeiras, estado da Paraíba, microrregião do Brejo Paraibano. Do banco de dados total, foram selecionadas 24 características independentes relacionadas à carcaça dos animais e duas variáveis dependentes (conformação e acabamento da carcaça) (Tabela 1).

Tabela 1. Estatísticas descritivas para os grupos de escores de conformação (2 e 3) e acabamento (2 e 3).

Variáveis independentes	Conformação				Acabamento			
	Escore 2		Escore 3		Escore 2		Escore 3	
	N = 37		N = 74		N = 48		N = 64	
	Média	DP	Média	DP	Média	DP	Média	DP
PCV (kg)	23.83	2.88	25.44	2.71	24.15	2.94	25.77	3.02
PVA (kg)	29.67	3.47	32.02	3.04	30.08	3.52	32.40	3.30
PCQ (kg)	13.93	1.42	15.12	1.50	14.21	1.50	15.28	1.72
PCF (kg)	13.67	1.29	14.87	1.44	13.96	1.40	15.01	1.67
RB (%)	58.69	3.60	59.56	3.70	59.06	3.99	59.41	3.40
RCQ (%)	46.31	3.49	46.52	3.16	46.63	3.58	46.39	2.99
RCF (%)	47.13	3.32	47.28	3.14	47.41	3.45	47.18	2.97
CEC (cm)	56.35	2.29	55.86	2.97	56.04	2.37	56.34	3.19
CIC (cm)	59.96	3.72	61.30	4.20	59.95	4.03	61.95	4.14
LG (cm)	16.92	3.12	17.49	2.68	16.82	3.14	18.02	2.90
LT (cm)	16.21	3.63	17.68	3.15	16.10	3.63	18.32	3.18
PG (cm)	54.27	7.29	55.20	4.22	53.88	4.00	55.96	6.38
PP (cm)	36.95	2.60	37.64	2.86	37.08	2.90	37.52	2.67
PT (cm)	67.26	4.12	68.43	2.96	66.87	2.55	69.08	3.66
PER (kg)	2.10	0.23	2.29	0.28	2.14	0.25	2.30	0.26
LOM (kg)	0.89	0.14	0.99	0.18	0.91	0.15	0.99	0.18
COS (kg)	2.05	0.28	2.17	0.30	2.10	0.29	2.14	0.30
PES (kg)	0.79	0.31	0.94	0.30	0.78	0.30	0.97	0.29
PAL (kg)	1.32	0.16	1.38	0.17	1.33	0.15	1.37	0.17
RPER (%)	30.85	3.28	30.87	2.92	30.79	3.09	30.76	3.05
RLOM (%)	13.07	2.09	13.27	1.72	13.04	1.96	13.22	1.89
RCOS (%)	30.02	3.70	29.23	3.10	30.17	3.51	28.61	3.19
RPES (%)	7.41	1.34	7.49	1.16	7.37	1.21	7.36	1.32
RPAL (%)	19.39	2.08	18.63	1.69	19.13	1.95	18.34	2.05

PCV - Peso de corpo vazio; PVA - Peso vivo ao abate; PCQ - Peso de carcaça quente; PCF - Peso de carcaça fria; RB - Rendimento biológico; RCQ - Rendimento de carcaça quente; RCF - Rendimento de carcaça fria; CEC - Comprimento externo da carcaça; CIC - Comprimento interno da carcaça; LG - Largura da garupa; LT - Largura do tórax; PG - Perímetro da garupa; PP - Perímetro do pernil; PT - Perímetro do tórax; PER - Pernil; LOM - Lombo; COS - Costela; PES - Pescoço; PAL - Paleta; RPER - Rendimento do pernil; RLOM - Rendimento do lombo; RCOS - Rendimento da costela; RPES - Rendimento do pescoço; RPAL - Rendimento da paleta.

Para a análise discriminante, os grupos foram formados em função dos escores (1 e 2) de conformação e acabamento das carcaças ovinas. O tamanho da amostra foi analisado de acordo com Hair Jr. *et al.* (2009), que recomendam uma proporção de 20 observações para cada variável preditora e que, no mínimo, o tamanho do menor grupo de uma categoria deve exceder o número de variáveis independentes.

A normalidade multivariada foi aplicada pelo teste de Mardia, que se baseia nos desvios de assimetria e curtose. As variáveis de carcaças ovinas não atenderam à premissa de normalidade multivariada ($P < 0,05$) (Tabela 2).

Tabela 2. Teste de normalidade multivariada Mardia para as variáveis independentes.

Teste estatístico	Valor	Probabilidade
Mardia assimetria	6058	<.0001
Mardia curtose	20.62	<.0001

Quando a suposição de normalidade multivariada não é atendida, métodos não-paramétricos podem ser usados para estimar as densidades específicas do grupo, a exemplo do teste adotado nesse estudo, k-vizinho mais próximo (K-NN) (ROSENBLATT, 1956). O K-NN é um algoritmo baseado em instância, que não requer requisitos de distribuição de dados para produzir um modelo eficaz (PAN *et al.*, 2020). Esse método assume que todas as instâncias correspondem a pontos no espaço n-dimensional (BARBON *et al.*, 2016) e, para classificar uma nova instância, o classificador K-NN seleciona k vizinhos mais próximos no conjunto de dados de acordo com uma distância métrica (TAHERI-GARAVAND *et al.*, 2019). Depois que k vizinhos são encontrados, um valor médio é calculado entre vizinhos e atribuídos com valor de previsão a uma instância desconhecida (BARBON *et al.*, 2017).

Para identificar a contribuição de cada variável medida e sua importância na separação dos grupos, a combinação linear de k variáveis chamada função discriminante ou função discriminante canônica foi empregada para a separação dos grupos. Sendo calculado pela seguinte fórmula:

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \dots + W_nX_{nk}$$

Em que:

Z_{jk} = score Z discriminante da função discriminante j para o objeto k

a = intercepto

W_i = peso discriminante para a variável independente i

X_{ik} = variável independente i para o objeto k

O procedimento *stepwise* foi utilizado para selecionar variáveis independentes com maior poder discriminatório sobre as variáveis dependentes. O procedimento *stepwise* é uma ferramenta popular de mineração de dados que usa significância estatística para selecionar as variáveis explicativas a serem usadas em um modelo matemático (SMITH, 2018), sendo o processo de seleção para adição ou remoção das variáveis foi realizado como base no teste estatístico Lambda de Wilks'.

A distância de Mahalanobis (D^2) foi utilizada para a separação dos grupos. A D^2 de Mahalanobis leva em consideração a correlação dos dados, pois é calculada usando o inverso

da matriz variância-covariância do conjunto de dados (MAESSCHALCK; JOUAN-RIMBAUD; MASSART, 2000). Os grupos foram declarados significativamente separados quando o valor de P mostrou-se menor que 0,05. A função da D^2 de Mahalanobis é dada por:

$$DM_{(\bar{X}_j, \bar{X}_k)} = \sqrt{(\bar{X}_j - \bar{X}_k)^T S^{-1} (\bar{X}_j - \bar{X}_k)}$$

Em que S^{-1} é a matriz de covariância inversa S entre \bar{X}_j e \bar{X}_k . Essa matriz é calculada usando a média ponderada das matrizes de covariância entre os grupos (OLATUNJI *et al.*, 2019). No KNN, a matriz covariância combinada é usada para calcular a distância de Mahalanobis.

As análises estatísticas foram realizadas por meio do programa estatístico SAS® OnDemand Academics. Para realizar o teste de normalidade multivariada de Mardia foi utilizado o procedimento %macro multnorm. O teste não-paramétrico K-NN foi realizado pelo método = *npar*. O procedimento STEPDISC, por intermédio do procedimento de seleção *stepwise* foi utilizado para encontrar um subconjunto de variáveis que melhor revele diferenças entre os grupos, ou seja, as variáveis com maior poder discriminatório. A D^2 de Mahalanobis foram calculadas usando o procedimento CANDISC, para realizar a análise discriminante canônica e encontrar combinações lineares das variáveis que melhor resume as diferenças entre grupos. E para calcular as funções discriminantes e classificar observações foi utilizado o procedimento DISCRIM.

3. RESULTADOS E DISCUSSÃO

Na Tabela 3 são apresentadas as variáveis independentes selecionadas pelo procedimento *stepwise*, para os escores de conformação (2 e 3) de carcaças ovinas e a D^2 de Mahalanobis.

Tabela 3. Variáveis independentes selecionadas pelo procedimento *stepwise* para os escores de conformação (2 e 3) de carcaças ovinas e a D^2 de Mahalanobis.

Número de variáveis	Variáveis selecionadas	Variáveis removidas	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	PCF	---	18.03	<.0001	0.858	<.0001
2	CEC	---	10.51	0.0016	0.782	<.0001
3	PES	---	4.10	0.0455	0.753	<.0001
D^2 de Mahalanobis mínimo entre os escores de conformação (2 e 3)						
Entre escores		D^2 mínimo		Pr > D^2		
2 e 3		1,448		<.0001		

PCF - Peso de carcaça fria; CEC - Comprimento externo da carcaça; PES - Pescoço.

No conjunto de variáveis independentes pode haver variáveis que pouco influenciam as variáveis dependentes e nesse sentido o procedimento *stepwise* é usado para selecionar quais variáveis com maior contribuição para o estudo (ALVES; LOTUFO; LOPES, 2013). De acordo Senra *et al.* (2007), esse procedimento baseia-se na observação de que algumas variáveis contribuem pouco para a eficiência média do modelo, sendo assim, uma vez identificadas podem ser retiradas.

Entre os dois grupos de escores de conformação avaliados (2 e 3), podemos observar diferença significativa ($P < 0,05$) com base na D^2 de Mahalanobis (Tabela 3). Um indivíduo é atribuído a um grupo específico caso sua pontuação discriminante for menor que o valor de corte, obtido pelo cálculo da distância média ponderada entre os centróides dos grupos (MARDIA; KENT; BIBBY, 2000). Então, os centróides dos grupos são calculados e, para cada indivíduo, as distâncias são avaliadas, sendo que um indivíduo é designado a um determinado grupo com base na menor distância dos centróides dos grupos. Com base nisso, podemos observar que a distância entre os grupos de escores de conformação avaliados foi de 1,448, resultando na diferenciação entre grupos. Os coeficientes canônicos das funções discriminantes são combinações lineares das variáveis originais em que os coeficientes maximizam a separação entre grupos (ALONZO; ROTH; ROBERTS, 2013). E essa diferença entre grupos avaliados indicam que esses respectivos grupos obtiveram comportamento diferentes, havendo separação entre os grupos de escore de conformação (2 e 3). No geral, das 24 variáveis originais, apenas três (PCF, CEC e PES) apresentaram maior poder de discriminação entre os grupos avaliados.

Na Tabela 4 são apresentados os coeficientes canônicos padronizados para a variável canônica (CAN1) referente às três variáveis mais discriminantes selecionadas pelo procedimento *stepwise* para os escores de conformação.

Tabela 4. Coeficientes canônicos padronizados para a variável canônica (CAN1) para os escores de conformação (2 e 3).

Variáveis selecionadas	Can1
PCF	1.087250546
CEC	-0.778586852
PES	0.415628072
Correlação Canônica	0,497
Variação total (%)	100

PCF - Peso de carcaça fria; CEC - Comprimento externo da carcaça; PES - Pescoço.

Os pesquisadores geralmente empregam coeficientes canônicos padronizados para ajudar na interpretação da contribuição de cada variável resposta, especialmente quando as variáveis não são proporcionais (ALKARKHI; ALQARAGHULI, 2020), sendo que a finalidade de utilização dos coeficientes canônicos é determinar as combinações lineares que fornecem a máxima diferenciação entre grupos. Nesse contexto, os coeficientes das combinações lineares são os coeficientes canônicos que indicam a contribuição parcial de cada variável original na composição do CAN (DIMAURO *et al.*, 2013).

Com as variáveis discriminantes identificadas e a função discriminante descrita, a próxima fase é a interpretação. Podemos observar que das três variáveis da função discriminante todas excedem o valor da carga discriminante preconizado por Hair Jr. *et al.* (2009) de, no mínimo, $\pm 0,40$, garantindo, assim, inclusão no processo de interpretação por serem consideradas robustas.

O CAN1, com correlação canônica 0,497, a característica que mais contribuiu para compor as combinações lineares foi a variável independente PCF. O resfriamento da carcaça tem por objetivo retardar ou evitar alterações microbianas, químicas e físicas, que reduzem a qualidade da carne. Esses fatores são dependentes da quantidade e distribuição da massa muscular e gordura presente na carcaça, que do ponto de vista biológico caracterizam a conformação. O fator confinamento e a aptidão para carne determinam maior deposição de gordura e conseqüentemente menor relação músculo:gordura (SANTOS *et al.*, 2010) e animais mais pesados. Quanto maior o peso da carcaça, melhor é a cobertura de gordura, sendo assim, o PCF é uma variável primordial para uma correta conservação e qualidade da carne e, conseqüentemente, sobre o valor final de comercialização.

A segunda variável independente a entrar no modelo foi o CEC. Esse resultado reforça o descrito por Caro *et al.* (2018) sobre a importância do comprimento da carcaça na variabilidade da conformação da carcaça. No estudo supracitado, a análise de componentes principais foi realizada utilizando variáveis de comprimento, largura e profundidade da carcaça. E o componente 1 foi responsável por 64% da variabilidade, sendo que os maiores escores corresponderam ao comprimento da carcaça e comprimento do lombo (escores $>0,40$). Por último, a variável independente PES também contribuiu para a variabilidade da conformação da carcaça. A raça Santa Inês é considerada um animal de grande porte e o pescoço é proporcional ao corpo. Por ser um animal com aptidão para corte e nesse estudo os animais estarem submetidos ao regime de confinamento, a quantidade e distribuição muscular e de gordura na base óssea podem ser mais elevados. Como descrito por Cezar e Sousa (2007b), os cortes de

pescoço, por terem desenvolvimento mais tardio, faz com que ocorra incremento de peso à medida que o peso da carcaça aumenta.

A análise discriminante também tem por finalidade atribuir observações a grupos previamente definidos. Como relatado por ElMasry *et al.* (2011), um conhecimento prévio para grupos predefinidos das amostras testadas é um pré-requisito primordial para a aplicação de análise discriminante, para diferenciação entre grupos. A Tabela 5 exhibe o resumo de quantas observações são classificadas corretamente ou incorretamente nos grupos de escores de conformação.

Tabela 5. Número de observações e percentual de classificação dos grupos de escores de conformação (2 e 3).

Escores de conformação	2	3	Total
2	37 100%	0 0.00%	37 100%
3	0 0.00%	74 100%	74 100%
Total	37 33.33%	74 66.67%	111 100%

Percebe-se que os grupos foram classificados corretamente em seu grupo de origem, indicando que a primeira função canônica apresentou bons resultados de discriminação para a separação dos grupos.

Na Tabela 6 são apresentadas as variáveis independentes selecionadas pelo procedimento *stepwise* e a D^2 de Mahalanobis para os escores de acabamento (2 e 3).

Tabela 6. Variáveis independentes selecionadas pelo procedimento *stepwise*, para os escores de acabamento (2 e 3) de carcaças ovinas e a D^2 de Mahalanobis.

Número de variáveis	Variáveis selecionadas	Variáveis removidas	F Value	Pr > F	Wilks' Lambda	Pr < Lambda
1	PVA	---	12.86	0.0005	0.895	0.0005
2	PT	---	6.31	0.0135	0.846	0.0001
3	CEC	---	4.58	0.0346	0.812	<.0001
D^2 de Mahalanobis mínimo entre os escores de acabamento (2 e 3)						
Entre escores 2 e 3		D^2 mínimo		Pr > D^2		
		0.929		<.0001		

PVA - Peso vivo ao abate; PT - Perímetro do tórax; CEC - Comprimento externo da carcaça.

Em relação aos escores de acabamento, a variável PVA foi a primeira a entrar no modelo discriminante, por ter maior diferença significativa entre grupos. A segunda variável a entrar no modelo é PT, e ocorre uma melhora na discriminação entre grupos, como evidencia pela diminuição no Lambda de Wilk's de 0,895 para 0,846. A adição da terceira variável CEC na função discriminante, melhorou a qualidade do modelo, como evidenciado pela diminuição no valor de Lambda de Wilk's (de 0,846 para 0,812).

Com isso, das 24 variáveis independentes propostas para o modelo geral dos escores de acabamento, três variáveis (PVA, PT e CEC) apresentaram maior poder de discriminação. Sendo assim, essas respectivas três variáveis apresentaram um poder discriminatório relevante no processo de estimação, constituindo a função discriminante. Para a D^2 de Mahalanobis (Tabela 6), podemos observar diferença estatística significativa ($P > 0,05$) entre grupos, sugerindo que os grupos de escores de acabamento (1 e 2) possuem comportamento diferentes.

Na Tabela 7 são apresentados os coeficientes canônicos padronizados para a CAN1, referente às três variáveis mais discriminantes selecionada pelo procedimento *stepwise* para os escores de acabamento.

Tabela 7. Coeficientes canônicos padronizados para a variável canônica (CAN1) para os escores de acabamento (2 e 3).

Variáveis selecionadas	Can1
PVA	0.8964425490
CEC	-.5730387197
PT	0.6793909749
Correlação Canônica	0,433
Variação total (%)	100

PVA - Peso vivo ao abate; PT - Perímetro do tórax; CEC - Comprimento externo da carcaça.

Assim como observado para os escores de conformação, percebe-se que as três variáveis inseridas na função discriminante excedem o valor da carga discriminante preconizado por Hair Jr. *et al.* (2009) de $\pm 0,40$, confirmando o bom ajuste do modelo.

O CAN1 apresentou correlação canônica de 0,433; a característica que mais contribuiu para compor as combinações lineares relacionadas ao acabamento foi a variável independente PVA. É de conhecimento amplo que o PVA tem potencial como indicador que afeta as características da carcaça e qualidade da carne. Como declarado por Jones *et al.* (2021), cordeiros mais pesados ao desmame têm probabilidade significativamente maior de apresentar carnes mais magras e musculosas. Porém, isso pode interferir na qualidade, uma vez que carcaças com baixa cobertura subcutânea pode causar encurtamento dos músculos pelo frio e,

consequentemente, menor maciez da carne. Isso reforça a importância desta variável independente na separação dos grupos relacionados à cobertura muscular.

As variáveis subsequentes a compor o modelo da função discriminante foram as independentes, PT e CEC. Como descrito por Restle *et al.* (2006), animais em estágio de terminação são preferíveis os mais compridos, que é o caso dos ovinos da raça Santa Inês e, como maior perímetro torácico. Em resumo, animais com essas características, consequentemente, são mais pesados e depositam maior quantidade de gordura de cobertura (ROSA *et al.*, 2014). Justificando a importância dessas variáveis independentes na variabilidade de acabamento da carcaça.

Assim como ocorre na análise de componentes principais, as variáveis canônicas são combinações lineares das variáveis originais, reduzindo a quantidade de informações redundantes e permitindo que as primeiras funções canônicas retenham o máximo de informações das variáveis originais. Nesse estudo foi possível observar que a primeira variável canônica (CAN1), tanto para os escores de conformação e acabamento (Tabela 4 e Tabela 6) envolveu uma variância total de 100%, isto é, um indicativo evidente de que o banco de dados analisado pode ser simplificado pela redução do espaço amostral, facilitando, consequentemente, a interpretação dos resultados.

Assim como observado para os escores de conformação, percebe-se também que os grupos de acabamento foram classificados corretamente em seu grupo de origem, indicando que a primeira função canônica apresentou bons resultados de discriminação para a separação dos grupos (Tabela 8).

Tabela 8. Número de observações e percentual de classificação dos grupos de escores de acabamento (2 e 3).

Escores de acabamento	2	3	Total
2	48 100%	0 0.00%	48 100%
3	0 0.00%	64 100%	64 100%
Total	48 42.86%	64 57.14%	112 100%

4. CONCLUSÃO

A análise discriminante multivariada mostrou-se eficiente para alocar os animais em seus grupos de origem (escores de carcaça). As variáveis com maior poder discriminatório para os escores de conformação carcaça foram peso de carcaça fria, comprimento externo da carcaça e

o pescoço e para acabamento de carcaça foram peso vivo ao abate e perímetro torácico. Esses resultados servirão de base para futuros estudos com o mesmo objetivo.

5. REFERÊNCIAS

- ALKARKHI, A. F.; ALQARAGHULI, W. A. A. Discriminant Analysis. In: Applied Statistics for Environmental Science with R. **Discriminant Analysis**. p. 173–190, 2020.
- ALONZO, M.; ROTH, K.; ROBERTS, D. Identifying santa barbara’s urban tree species from AVIRIS imagery using canonical discriminant analysis. **Remote Sensing Letters**, v. 4, n. 5, p. 513–521, 2013.
- ALVES, M. F.; LOTUFO, A. D. P.; LOPES, M. L. M. Seleção de variáveis stepwise aplicadas em redes neurais artificiais para previsão de demanda de cargas elétricas. **Proceeding Series of the Brazilian Society of Applied and Computational Mathematics**, v. 1, n. 1, p. 1–6, 2013.
- BARBON, A. P. A. C. *et al.* Storage time prediction of pork by Computational Intelligence. **Computers and Electronics in Agriculture**, v. 127, p. 368–375, 2016.
- BARBON, A. P. A. DA C. *et al.* Development of a flexible Computer Vision System for marbling classification. **Computers and Electronics in Agriculture**, v. 142, p. 536–544, 2017.
- CARO, I. *et al.* Conformation characteristics of suckling lambs carcasses from the Spanish local breeds Churra and Castellana and the non-native breed Assaf determined using digital photographs. **Small Ruminant Research**, v. 160, p. 89–94, 2018.
- CEZAR, M. F.; SOUSA, A. H. DE. **Carcaças Ovinas e Caprinas**. 1.ed. João Pessoa: UFCG, 2007.
- DIMAURO, C. *et al.* Use of the canonical discriminant analysis to select SNP markers for bovine breed assignment and traceability purposes. **Animal Genetics**, v. 44, n. 4, p. 377–382, 2013.
- ELMASRY, G. *et al.* Quality classification of cooked , sliced turkey hams using NIR hyperspectral imaging system. **Journal of Food Engineering**, v. 103, n. 3, p. 333–344, 2011.
- HAIR JR, J. *et al.* **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009.
- JEON, H. *et al.* Discrimination of origin of sesame oils using fatty acid and lignan profiles in combination with canonical discriminant analysis. **JAACS, Journal of the American Oil Chemists’ Society**, v. 90, n. 3, p. 337–347, 2013.
- JONES, A. G. *et al.* Using a lamb’s early-life liveweight as a predictor of carcass quality. **Animal**, v. 15, n. 1, 2021.
- MAESSCHALCK, R.; JOUAN-RIMBAUD, D.; MASSART, D. L. The Mahalanobis distance. **Chemometrics and Intelligent Laboratory Systems**, v. 50, n. 1, p. 1–18, 2000.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate Analysis**. London: Academic Press, 2000.
- OLATUNJI, O. O. *et al.* Property-based biomass feedstock grading using k-Nearest Neighbour technique. **Energy**, 2019.
- PAN, Y. *et al.* A new fast search algorithm for exact k-nearest neighbors based on optimal triangle-inequality-based check strategy. **Knowledge-Based Systems**, v. 189, p. 1–11, 2020.
- RESTLE, J. *et al.* Silagem de diferentes híbridos de milho para produção de novilhos

superjovens. **Revista Brasileira de Zootecnia**, v. 35, n. 5, p. 2066–2076, 2006.

ROSA, B. L. *et al.* Correlações entre medidas corporais e características das carcaças de tourinhos Nelore terminados em confinamento. **Boletim de Indústria Animal**, v. 71, n. 4, p. 371–380, 2014.

ROSENBLATT, M. Remarks on some nonparametric estimates of a density function. **Annals of Mathematical Statistics**, v. 27, p. 832–837, 1956.

SANTOS, J. R. . *et al.* Efeito da suplementação na composição física e centesimal da paleta, do costilhar e do pescoço de cordeiros Santa Inês terminados em pastejo. **Revista Brasileira de Medicina Veterinária e Zootecnia**, v. 62, n. 4, p. 906–913, 2010.

SAS - INSTITUTE INCORPORATION. **SAS® Studio**, 2021.

SENRA, L. F. A. DE C. *et al.* Estudo sobre métodos de seleção de variáveis em DEA. **Pesquisa Operacional**, v. 27, n. 2, p. 191–207, 2007.

SMITH, G. Step away from stepwise. **Journal of Big Data**, v. 5, n. 1, p. 2–12, 2018.

TAHERI-GARAVAND, A. *et al.* Real-time nondestructive monitoring of Common Carp Fish freshness using robust vision-based intelligent modeling approaches. **Computers and Electronics in Agriculture**, v. 159, p. 16–27, 2019.