

UNIVERSIDADE FEDERAL DE VIÇOSA

LAYLA CRISTIEN DE CÁSSIA MIRANDA DIAS

ANÁLISE MULTIVARIADA NA AVALIAÇÃO DE CARÇAÇA DE SUÍNOS

VIÇOSA – MINAS GERAIS

2020

LAYLA CRISTIEN DE CÁSSIA MIRANDA DIAS

ANÁLISE MULTIVARIADA NA AVALIAÇÃO DE CARCAÇA DE SUÍNOS.

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Zootecnia, para obtenção do título de *Magister Scientiae*.

Orientadora: Renata Veroneze

VIÇOSA – MINAS GERAIS

2020

Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa

T

D541a Dias, Layla Cristien de Cássia Miranda, 1995-
2020 Análise multivariada na avaliação de carcaça de suínos /
Layla Cristien de Cássia Miranda Dias. – Viçosa, MG, 2020.
45 f. : il. (algumas color.) ; 29 cm.

Inclui anexo.

Orientador: Renata Veroneze.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 40-43.

1. Suínos - Carcaças. 2. Análise multivariada. 3. Análise de componentes principais. 4. Modelos lineares (Estatística).
I. Universidade Federal de Viçosa. Departamento de Zootecnia. Programa de Pós-Graduação em Zootecnia. II. Título.

CDD 22. ed. 636.40821

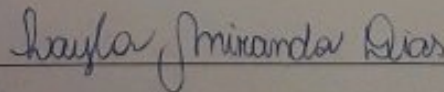
LAYLA CRISTIEN DE CÁSSIA MIRANDA DIAS

ANÁLISE MULTIVARIADA NA AVALIAÇÃO DE CARÇA DE SUÍNOS

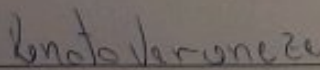
Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Zootecnia, para obtenção do título de *Magister Scientiae*.

APROVADA: 28 de julho de 2020.

Assentimento:



Layla Cristien de Cássia Miranda Dias
Autora



Renata Veroneze
Orientadora

Às meninas da minha vida, mami e Manu

AGRADECIMENTOS

Agradeço primeiramente a Deus, por tudo que me foi concedido até então. Pela minha vida, saúde e por me guiar em mais esta etapa da minha vida.

Aos meus pais, Rita e Maciel, por todo o apoio, todo carinho, amor, paciência e por sempre acreditarem em mim.

À minha irmã, Emanuelle, por ser minha companheira de vida e melhor amiga e, ao seu próprio jeito, me apoiar sempre e nunca me deixar desistir.

À minha orientadora, Renata Veroneze, por ter me conduzido durante a execução deste trabalho, pela paciência, confiança e tempo em mim depositados.

Ao Professor Paulo Sávio, por todo incentivo e paciência.

Aos meus amigos, que sempre estiveram por perto (mesmo que distantes), por todo suporte, por me aturarem sempre,

A todos aqueles que estiveram, de alguma forma, envolvidos no meu crescimento pessoal e profissional.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“Não é preciso ter olhos abertos para ver o sol, nem é preciso ter ouvidos afiados para ouvir o trovão. Para ser vitorioso você precisa ver o que não está visível”. (Sun Tzu)

RESUMO

DIAS, Layla Cristien de Cássia Miranda, M.Sc., Universidade Federal de Viçosa, julho de 2020. **Análise multivariada na avaliação de carcaça de suínos.** Orientadora: Renata Veroneze.

As técnicas multivariadas podem ser aplicadas a dados de carcaça e de cortes de suínos medidos em plantas frigoríficas, visando a identificação das variáveis mais relevantes no conjunto de dados e, portanto, contribuindo para otimização da fenotipagem. Objetiva-se identificar as variáveis mais relevantes relacionadas aos cortes comerciais de carcaça através da análise de componentes principais e avaliar a eficiência da análise discriminante para a classificação de duas linhas macho com diferentes idades ao abate utilizando dados de carcaça e cortes comerciais de carne suína. O experimento foi conduzido em uma granja comercial de ciclo completo, localizada na zona rural do município de Ponte Nova do Estado de Minas Gerais. Foi utilizado sêmen de reprodutores de duas linhas macho distintas (L1 e L2) em fêmeas Camborough. Foram realizadas três inseminações por fêmea, utilizando sempre machos da mesma linha (L1 ou L2), totalizando 167 coberturas experimentais direcionadas durante 11 dias. Os animais foram identificados após o nascimento por meio de tatuagem nas duas orelhas e foram criados seguindo o sistema de produção da granja, utilizando mesma dieta e manejo para todos os animais. O abate de animais foi realizado no Frigorífico Industrial Vale do Piranga S/A localizado no município de Oratórios do Estado de Minas Gerais. Foram abatidos 433 animais por amostragem, sendo 213 progênies da L1 e 220 da L2, com idades variando entre 145 a 191 dias. Os animais foram divididos em dois grupos de acordo com a idade em que foram abatidos: animais abatidos entre 145 e 170 dias (1) e entre 185 e 191 dias (2). Sendo assim, foram formados quatro grupos ($k = 4$), representado pela combinação das duas linhagens (L1 e L2) e de duas idades de abates (1 e 2). Foram mensuradas as variáveis espessura de toucinho (ET), espessura de músculo (EM), percentual de carne magra (CM), peso de carcaça quente automático (PCQAut), peso de carcaça quente (PCQ), peso da carcaça resfriada (PCResf), peso de pernil para presunto (PerPresunto), peso de pernil para picanha (PerPicanha), peso de pernil para alcatra (PerAlcatra), peso do pernil para bacon (PerBacon), peso do pernil inteiro (Pernil), peso da paleta para presunto (PaPresunto), peso da paleta para bacon (PaBacon), peso da paleta inteira (Paleta), peso do lombo (Lombo), peso do toucinho do carré (toucinho), peso do carré (Carre), peso da barriga (barriga), Peso da barriga e costela inteiro (BarCost), peso da costela (costela), peso da fraldinha (fraldinha), peso da sobrepaleta inteira (sobrepaleta), peso do

filé limpo (FileLimpo), peso do filé (File) e peso da papada (Papada). Os dados foram processados e submetidos as análises de componentes principais e discriminante utilizando o software R. Os dois primeiros componentes foram capazes de explicar 67,28% da variação total dos dados, sendo necessários pelo menos 9 componentes para explicar mais de 90% da variabilidade. Somente 5 componentes principais apresentaram autovalores maiores que 0,7, assim as variáveis com maior valor absoluto dentro desses componentes foram identificadas como as variáveis mais relevantes nesse conjunto de dados originais estudadas (PCResf, CM, PaPresunto, EM e Papada. A análise visual do gráfico de scores em relação aos componentes principais 1 e 2, revela que não ocorreu a formação de clusters e que não é possível, a partir desses componentes, separar os indivíduos pertencentes aos quatro grupos avaliados (combinação de idade de abate e linha). A análise discriminante foi realizada utilizando somente as cinco variáveis identificadas como mais importantes na análise de componentes principais. Utilizando populações de treinamento e validação verificou-se que de modo geral o percentual de acertos de classificação utilizando a análise discriminante foi maior para a idade de abate mais jovem (145-170 dias), conseqüentemente, as maiores taxas de erro foram observadas para os animais com maiores pesos ao abate. A coleta de dados das variáveis PCResf, CM, PaPresunto, EM e Papada em frigoríficos é recomendada, permitindo, portanto, a otimização da fenotipagem apesar de não possibilitar a diferenciação entre os grupos avaliados. Além disso, uma grande taxa de erro foi observada na classificação das amostras, sendo mais difícil classificar animais abatidos mais velhos.

Palavras-chave: Suínos – Carcaças. Análise multivariada. Análise de componentes principais. Modelos lineares (Estatística).

ABSTRACT

DIAS, Layla Cristien de Cássia Miranda, M.Sc., Universidade Federal de Viçosa, July, 2020. **Multivariate analysis in the evaluation of pig carcasses.** Adviser: Renata Veroneze.

Multivariate techniques may be applied to carcass and pig meat cuts data measured in slaughterhouses for the identification of the most relevant variables in the data set and, therefore, contributing to the phenotyping optimization. Therefore, we aimed to apply principal components for the identification of the most relevant carcass and pork cut variables measured at slaughter and to evaluate the efficiency of linear discriminant analysis on the classification of two male lines slaughtered at different ages using carcass and meat cut data. The trial was conducted in a commercial pig farm located at Ponte Nova, Minas Gerais state. Semen from two male lines (L1 and L2) was used in the insemination of Camborough sows. Each sow was inseminated three times using the same male line. In total 167 sows were inseminated in eleven days. After birth, the piglets were tattooed on the ear and raised following the pig farm protocols. All animals received the same diet and management. The animals were slaughtered at Frigorífico Industrial Vale do Piranga S/A, located at Oratórios, Minas Gerais state. A sample of 433 pigs with age ranging from 145 to 191 days were slaughtered, being 213 progenies of L1 and 220 progenies of L2. The animals were split in two groups according to the slaughter age: group 1 from 145 to 170 days and group 2 from 185 to 191 days. Thus, four groups were considered in the analysis, which correspond to the combination of male line (L1 and L2) and slaughter age. The variables measured were backfat thickness (ET), muscle depth (EM), percentage of lean meat (CM), automatic carcass weight (PCQAut), hot carcass weight (PCQ), chilled carcass weight (PCResf), pork leg weight for ham (PerPresunto), pork rump cap weight (PerPicanha), pork rump steak (PerAlcatra), pork leg weight for bacon (PerBacon), pork leg weight (Pernil), shoulder weight for ham (PaPresunto), shoulder weight for bacon (PaBacon), shoulder weight (Paleta), loin weight (Lombo), loin backfat weight (toucinho), pork rib chop weight (Carre), belly weight (barriga), rib and belly weight (BarCost), rib weight (costela), pork flank steak (fraldinha), pork boston butt weight (sobrepaleta), tenderloin weight without fat (FileLimpo), tenderloin weight (File) and jowl weight (Papada). All analysis was done using R software. The first and second principal components explained 67,28% of the total data variability, being necessary at least nine components to explain more than 90% of the

variability. Only five principal components presented eigenvalues higher than 0.7, thus variables with the highest absolute coefficients on these components were identified as the most important for this dataset (PCResf, CM, PaPresunto, EM e Papada). Visual inspection of the first and second principal components score plot did not allow the identification of clusters. Therefore, using these components is not possible to separate individuals from the four groups evaluated (combination of line and age). Discriminant analysis was done using only the five most important variables identified in the principal components analysis. Using training and validation populations it was verified that the classification hits were higher for the young slaughter age group (145-170 days) than for the age group two. Thus, highest misclassification was observed for older animals, which have higher slaughter weight. The slaughterhouse should register the variables PCResf, CM, PaPresunto, EM e Papada. Although those variables did not allow group identification they will optimize the phenotyping. High error rate was observed in sample classification according to line and slaughter age group, being more difficult to classify older animals.

Keywords: Pigs – Carcasses. Multivariate analysis. Principal component analysis. Linear models (Statistics).

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 Heatmap – correlações entre as características..... | 28 |
| Figura 2 Scree-plot. Gráfico que demonstra o número de componentes a serem escolhidos na análise..... | 29 |
| Figura 3 Gráfico de cargas das variáveis considerando os dois primeiros componentes principais..... | 32 |
| Figura 4 Gráfico de scores dos dois primeiros componentes principais para os quatro grupos avaliados..... | 33 |
| Figura 5 Gráfico de scores individuais da análise discriminante..... | 36 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 Identificação e número de observações dos grupos avaliados..... | 20 |
| Tabela 2 Descrição das características mensuradas durante o abate dos animais e processamento dos cortes comerciais..... | 21 |
| Tabela 3 Descrição das características..... | 27 |
| Tabela 4 Componentes principais e percentual explicado por cada componente..... | 30 |
| Tabela 5 Coeficientes de ponderação dos Componentes principais de menor variância para as características PCRsf, CM, PaPresunto, EM e Papada. | 31 |
| Tabela 6 Coeficientes da análise linear discriminante (LDA)..... | 34 |
| Tabela 7 Classificação dos indivíduos nos grupos em relação a características de carcaça..... | 35 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|------------|---|
| ABCS | Associação Brasileira dos Criadores de Suínos |
| ABPA | Associação Brasileira de Proteína Animal |
| BarCost | Peso da barriga e costela inteiro |
| Barriga | Peso da barriga |
| Carre | Peso do carré inteiro |
| CM | Carne magra |
| Costela | Peso da costela |
| EM | Espessura de músculo |
| ET | Espessura de toucinho |
| File | Peso de filé inteiro |
| FileLimpo | Peso do filé limpo e inteiro |
| Fraldinha | Peso da fraldinha |
| LDA | Análise Linear Discriminante |
| Lombo | Peso do carré e lombo |
| PaBacon | Peso da paleta e bacon |
| Paleta | Peso da paleta inteira |
| Papada | Peso de papada |
| PaPresunto | Peso da paleta para presunto |
| PCA | Análise de Componentes Principais |

| | |
|-------------|-----------------------------------|
| PCQ | Peso de carcaça quente |
| PCQAut | Peso de carcaça quente automático |
| PCResf | Peso da carcaça resfriada |
| PePresunto | Peso do pernil para presunto |
| PerAlcatra | Peso do pernil para alcatra |
| PerBacon | Peso do pernil para bacon |
| Pernil | Peso do pernil inteira |
| PerPicanha | Peso do pernil para picanha |
| Sobrepaleta | Peso da sobrepaleta inteira |
| Toucinho | Peso do carré e toucinho |

SUMÁRIO

| | |
|---|----|
| 1. INTRODUÇÃO | 15 |
| 2. REVISÃO DE LITERATURA | 16 |
| 3. MATERIAL E MÉTODOS | 19 |
| 3.1. <i>Descrição dos dados</i> | 19 |
| 3.2. <i>Imputação</i> | 23 |
| 3.3. <i>Análise estatística</i> | 23 |
| 3.4. <i>Análise de componentes principais (PCA)</i> | 24 |
| 3.5. <i>Análise discriminante (LDA)</i> | 25 |
| 3.6. <i>Validação cruzada</i> | 25 |
| 4. RESULTADOS | 26 |
| 4.1. <i>Estatística descritiva e correlação</i> | 26 |
| 4.2. <i>Análise de componentes principais (PCA)</i> | 28 |
| 4.3. <i>Análise discriminante (LDA)</i> | 34 |
| 4 DISCUSSÃO | 37 |
| 5 CONCLUSÃO | 39 |

1. INTRODUÇÃO

Os primeiros suínos foram trazidos para o Brasil em 1532 (ABCS) e ao longo do tempo a suinocultura tornou-se uma atividade muito difundida no país, com grande tecnificação e importância econômica. Atualmente, produtos cárneos de origem suína têm grande participação no mercado mundial, no Brasil foram produzidas 3.979 mil toneladas de carne suína em 2018, sendo 16% dessa produção destinada à exportação, o que conferiu ao Brasil a quarta colocação no ranking mundial no mercado de carne suína (ABPA). Os cortes de carne suína são os principais produtos de exportação do Brasil (ABPA) o que torna evidente a importância de pesquisas utilizando dados de frigoríficos e que avaliem informações de carcaça e de peso e rendimento dos cortes de maior importância comercial para os mercados interno e externo.

O melhoramento animal proporcionou avanços consideráveis em diversas características de interesse econômico como a redução da espessura de toucinho e conversão alimentar e aumento do tamanho de leitegada e ganho de peso (MERKS; VRIES, 2002). A caracterização do material genético disponível, o acompanhamento do progresso genético das linhas, e o planejamento para a coleta e incorporação de novos fenótipos na avaliação genética são fundamentais para que o melhoramento de suínos continue sendo bem sucedido. Contudo, ainda existe pouca ênfase em características relacionadas a qualidade da carcaça e rendimento de cortes e produtos no frigorífico. Entender as diferenças entre as carcaças dos animais a serem abatidos é um importante instrumento na produção animal, influenciando o destino e o valor comercial do produto (GERALDO et al., 2018). Atualmente, existem diversas linhas genéticas no mercado brasileiro, de diferentes empresas ou da mesma empresa, sendo economicamente relevante entender as particularidades das linhas quanto a qualidade da carcaça.

A indústria frigorífica comercializa uma grande quantidade de cortes e produtos suíno, assim a coleta de dados provenientes da indústria resulta em banco de dados com grande número de variáveis, sendo que muitas delas são altamente correlacionadas. As técnicas multivariadas de análise de dados, são técnicas estatísticas para análise de dados envolvendo duas ou mais variáveis em uma amostra (LOESCH; HOELTGEBAUM, 2017). Essas técnicas ganham destaque na análise de dados com grande número de variáveis, pois permitem a redução de dimensionalidade do banco de dados e o agrupamento de informações semelhantes entre si (NGUYEN; HOLMES, 2019), além

de aperfeiçoarem a compreensão dos dados possibilitando um melhor entendimento, mais completo e realista do conjunto de informações (HENRIQUES, 2018).

Dentre as técnicas multivariadas, a análise discriminante (LDA) é comumente utilizada para classificar e distinguir grupos, permitindo separar indivíduos de uma mesma população em dois ou mais grupos (KHATTREE; NAIK, 2000), a classificação dos indivíduos é feita por meio de um conjunto de regras que podem ser utilizadas para a classificação de novos indivíduos não envolvidos no processo inicial (JOHNSON; WICHERN, 2007). Já a análise de componentes principais (PCA), também considerada como uma importante técnica de análise multivariada, é utilizada para reduzir a complexidade de dados de alta dimensionalidade, facilitando a interpretação e identificação de padrões. Além disso, a PCA pode ser utilizada para a identificação de variáveis de maior importância, promovendo a redução de dimensionalidade com a menor perda possível de informações (KHERIF; LATYPOVA, 2016), também pode ser utilizado para reconhecimento de grupos através da identificação de características responsáveis por essa diferenciação (ALKARKHI; ALQARAGHULI, 2020).

As técnicas multivariadas podem ser aplicadas a dados de carcaça e de cortes de suínos medidos em plantas frigoríficas, visando a identificação das variáveis mais relevantes no conjunto de dados e a possibilidade da identificação de linhas a partir dos cortes obtidos. Sendo assim, objetiva-se identificar as variáveis mais relevantes relacionadas aos cortes comerciais de carcaça através da análise de componentes principais e avaliar a eficiência da análise discriminante para a classificação de duas linhas macho com diferentes idades ao abate utilizando dados de carcaça e cortes comerciais de carne suína.

2. REVISÃO DE LITERATURA

Melhoramento de suínos

A domesticação dos suínos, assim como a de outros animais começou quando os homens deixaram de ser nômades e formaram aldeias, a partir de então a carne suína passou a ser a principal fonte de alimentação dos aldeões (CARDOSO; OYAMADA; SILVA, 2015). Os primeiros suínos no continente Americano foram introduzidos por Cristóvão Colombo, em 1493. No entanto, no Brasil, os animais foram inseridos apenas em 1532, nos primórdios da colonização brasileira, por Martim Afonso de Souza, a fim de que os portugueses presentes em território brasileiro pudessem

manter o padrão de alimentação que possuíam na Europa (ABCS, 2014) e, desde então, a carne suína passou a ocupar um importante papel na alimentação da população brasileira.

O melhoramento genético de suínos no Brasil teve início a partir da construção, pela Associação Brasileira de Criadores de Suínos (ABCS), das Estações de testes de Reprodutores de Suínos (ETRS) na década de 70 (LOPES et al., 1998). Inicialmente, o melhoramento genético de suínos focou em características relacionadas ao ganho de peso, conversão alimentar e espessura de toucinho (SUZUKI et al., 2005). Contudo, como a escolha das características a serem melhoradas se apoia nas exigências dos produtores de suínos e mercado consumidor, houve mudanças nos objetivos de seleção ao longo dos anos e um número cada vez maior de características passaram a ser incorporadas aos programas de melhoramento, como por exemplo, fenótipos relacionados a robustez, bem estar e valor do produto final (CARDOSO; OYAMADA; SILVA, 2015).

O mercado de produtos suínos exige animais com maior deposição de carne magra e alto potencial de crescimento (BOROSKY et al., 2011) e eficiência alimentar, além de características relacionadas à qualidade de carne, espessura de toucinho e gordura intramuscular (ČANDEK-POTOKAR; MONIN; ŽLENDER, 2002). Merks et al. (2012) reportaram que os programas de melhoramento iriam incluir como objetivos de seleção vitalidade e uniformidade do nascimento ao abate, redução de uso de antibióticos e aumento da segurança dos alimentos. Harlizius et al. (2020) destacaram a importância e as oportunidades advindas da seleção para maior resiliência e apontaram a variabilidade em dados de performance, como consumo, e reprodutivos como possíveis fenótipos para resiliência. Sendo assim, a relação de características relacionadas à importância econômica e social com potencial para inclusão nos programas de melhoramento cresce continuamente e tem sido potencializada pelo surgimento de novas formas de mensuração de fenótipos (HERMESCH et al., 2015) em larga escala.

O uso de novas tecnologias para fenotipagem dos animais como a utilização de imagens para mensurar características de carcaça (ALMEIDA et al., 2019), balanças e comedouros automáticos, sensores, entre outros, faz com que haja um crescente aumento no volume de dados produzido, impondo aos programas de melhoramento genético o desafio de manipular e analisar bancos de dados cada vez maiores e mais complexos.

Análise multivariada

A alta dimensionalidade dos bancos de dados utilizados nas mais diversas áreas de estudo gera dificuldades de processamento e armazenamento de dados, na complexidade de análise e aumenta a demanda de tempo e recursos computacionais (HAN et al., 2018). Dessa forma, extrair as variáveis mais representativas dentro do banco de dados pode ser uma alternativa ao uso dos dados originais (DORNAIKA; KHODER, 2020). O uso de técnicas multivariadas permite que sejam consideradas diversas variáveis para cada indivíduo em estudo, de forma que o relacionamento entre as variáveis e o comportamento real dessas variáveis seja considerado na análise (ALKARKHI; ALQARAGHULI, 2020a), além da redução da dimensionalidade do banco de dados (FERREIRA et al., 2003; CHOWDHURY et al., 2020; LEE et al., 2019).

A escolha da técnica multivariada a ser utilizada, ou de uma combinação delas é determinada pela natureza dos dados a serem avaliados. Dentre as diversas técnicas existentes, a análise de componentes principais (PCA) é amplamente difundida, sendo uma importante ferramenta para a avaliação da variabilidade total de um conjunto de variáveis e eliminação da redundância. A técnica permite a redução da dimensionalidade (OLIVEIRA et al., 2020) por meio de transformação ortogonal das variáveis em um pequeno grupo de variáveis não correlacionadas, mas que explicam a maior parte da variabilidade existente no conjunto de dados original (KHERIF; LATYPOVA, 2016).

A análise discriminante linear (LDA), a qual é empregada, dentre outras circunstâncias, na redução da dimensionalidade do banco de dados (YEH; WANG, 2013) e na determinação da contribuição de diversas variáveis para a distinção de grupos, possibilitando o reconhecimento de padrões existentes (GARRIDO-FERNÁNDEZ; LEÓN-CAMACHO, 2019), sendo que nessa abordagem os grupos envolvidos no estudo são previamente conhecidos (ALKARKHI; ALQARAGHULI, 2019). Portanto, a análise linear discriminante tem em vista a discriminação entre dois ou mais grupos objetivando a posterior classificação de novos indivíduos (REGAZZI; CRUZ, 2020).