

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

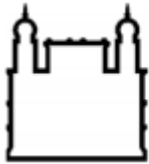
**INSTITUTO CARLOS CHAGAS
MESTRADO EM BIOCÊNCIAS E BIOTECNOLOGIA**

ELOIZA KAUANNA GONÇALVES DIAS FERREIRA

**USO DE FERRAMENTAS DE *DATA SCIENCE* PARA AVALIAÇÃO DOS DADOS
ZOTÉCNICOS DE PRODUÇÃO DE LINHAGENS DE CAMUNDONGOS: UMA
PROPOSTA DE MELHORIAS AO GERENCIAMENTO DO PLANTEL**

CURITIBA

2021



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO CARLOS CHAGAS
Mestrado em Biociências e Biotecnologia

ELOIZA KAUANNA GONÇALVES DIAS FERREIRA

USO DE FERRAMENTAS DE *DATA SCIENCE* PARA AVALIAÇÃO DOS DADOS
ZOOTÉCNICOS DE PRODUÇÃO DE LINHAGENS DE CAMUNDONGOS: UMA
PROPOSTA DE MELHORIAS AO GERENCIAMENTO DO PLANTEL.

Dissertação apresentada como parte dos requisitos
para obtenção do título de Mestre pelo Programa de
Pós-Graduação em Biociências e Biotecnologia.

Orientador: Guilherme Ferreira Silveira

Coorientador: Giovanny A. C. A. Mazzarotto

Curitiba

2021

Dedico esse trabalho a todos que me incentivaram e me apoiaram nesse desafio.

AGRADECIMENTOS

Agradeço a todos que ajudaram na construção desse trabalho. Primeiramente, devo agradecer ao meu orientador Guilherme Ferreira Silveira, pelo apoio e suporte durante o período em que trabalhamos juntos. Agradeço por todo ensinamento, paciência e dedicação.

Agradeço ao meu coorientador Giovanny Augusto Camacho Antevere Mazzarotto, pelo acompanhamento do trabalho e ensinamento sobre criação de camundongos em biotério.

Agradeço ao grupo do Laboratório de Criação e Experimentação Animal do Instituto Carlos Chagas, por me acolherem e me acompanharem na rotina do biotério, principalmente a Tania Schepainski do Rocio, a quem tive maior contato e me apresentou as funcionalidades do software BioterC.

Agradeço ao meu grupo de pesquisa DataLab, por me acompanharem e estarem presentes durante a finalização desse trabalho.

Agradeço ao Programa de Pós-graduação de Biociências e Biotecnologia do Instituto Carlos Chagas – Fiocruz/PR pelo apoio e a bolsa de estudos para desenvolver o mestrado.

Agradeço minha família, por todo o apoio e torcida durante essa trajetória.

“O círculo agora está completo. Quando o deixei, eu era só o aprendiz; agora eu sou o mestre.” (Darth Vader – Uma nova esperança)

RESUMO

O uso de animais para fins científicos é prática comum desde o século XVII. Com a promulgação da Lei 11794/2008 (Lei Arouca), posteriormente regulamentada pelo Decreto 6.899/2009, iniciou-se uma corrida para a implementação de mudanças no cenário da experimentação animal no Brasil, com ênfase na necessidade de melhorias em seu gerenciamento. Com o objetivo de melhorar a capacidade do processo gerencial do biotério do Instituto Carlos Chagas, foi desenvolvido e implementado em 2014 o *software* BioterC. Após 5 anos de uso ininterrupto desse *software*, o presente trabalho buscou compilar e analisar informações de seu banco de dados utilizando técnicas de Análise Exploratória de Dados (AED), a fim de responder perguntas referentes ao manejo da colônia e do biotério como um todo, identificando padrões de criação de camundongos. As análises foram realizadas utilizando ferramentas de mineração de dados, área do conhecimento que faz parte da Ciência de Dados (*Data Science*), em linguagem de programação *Python*. Os resultados obtidos demonstraram que, de setembro de 2014 até junho de 2019 foram produzidos 15.106 animais (1.808 entradas). Destes, a linhagem C57BL/6 (3.553 nascimentos), Balb/c (3.452 nascimentos) e Swiss (3.025 nascimentos) foram as mais produzidas. O ano de 2018 foi o que obteve maior produção de animais entre a maioria das linhagens. A distribuição de nascimentos/parto para o grupo isogênico apresentou uma média de 6-7 nascimento/parto e para o grupo heterogênico de 10 nascimento/parto. A média de dias do intervalo entre partos (IEP) foi de aproximadamente 35 dias. Já os registros do período entre o nascimento e o desmame apresentou média de 18,97 dias e mediana de 19 dias. Quanto os dados de fornecimento de camundongos, foram apresentados 403 registros de solicitações de animais de laboratórios internos e externos. Para os dados de descarte, o principal motivo identificado foi o Descarte Zootécnico, apresentando total de 2.286 camundongos. O presente trabalho proporcionou a identificação de padrões de criação de camundongos, corroborando com nossa hipótese de que seria possível identificar padrões de criação animal através de ferramenta de Ciência de Dados, além de provocar melhorias no manejo e processo de criação animal do LACEA, desde o nascimento até o descarte.

Palavras-Chave: Análise Exploratória dos Dados; Biotério; Animais de Laboratório; Mineração de Dados; *Python*.

ABSTRACT

The use of animals for scientific purposes has been a common practice since the 17th century. With the enactment of Law 11794/2008 (Lei Arouca), subsequently regulated by Decree 6.899 / 2009, a race began to implement changes in the scenario of animal experimentation in Brazil, with an emphasis on the need for improvements in its management. In order to improve the capacity of the management process of the animal facility at the Carlos Chagas Institute, the BioterC software was developed and implemented in 2014. After 5 years of uninterrupted use of this software, the present work sought to compile and analyze information from its database using Exploratory Data Analysis (AED) techniques, in order to answer questions regarding the management of the colony and the animal facility as a whole, identifying mouse breeding patterns. The analyzes were performed using data mining tools, an area of knowledge that is part of Data Science (Data Science), in Python programming language. The results obtained showed that, from September 2014 to June 2019, 15,106 animals (1,808 entries) were produced. Of these, the C57BL / 6 (3,553 births), Balb / c (3,452 births) and Swiss (3,025 births) lineages were the most produced. 2018 was the year with the highest animal production among most strains. The distribution of births / births for the isogenic group showed an average of 6-7 per birth and for the heterogeneous group of 10 births / births. The average number of days between the birth intervals (IEP) was approximately 35 days. The records of the period from birth to weaning showed an average of 18.97 days and a median of 19 days. Regarding the data on the supply of mice, 403 records of animal requests were submitted between internal and external laboratories. For the disposal data, the Zootechnical Disposal was the main reason, with a total of 2,286 mice. The present work provided the identification of mouse breeding patterns, corroborating our hypothesis that it would be possible to identify animal breeding patterns through a Data Science tool, in addition to causing improvements in LACEA's management and animal breeding process, since the beginning of the year, birth to disposal.

Keywords: Exploratory Data Analysis (AED); Animal Facility; Laboratory Animals; Data Mining; Python.

LISTA DE FIGURAS

Figura 1.1 - Diagrama de conhecimentos que agregados dão origem à área de Ciência de Dados. Fonte: Medium, 2018.	8
Figura 3.1 - Fluxograma do processamento de dados do BioterC.	11
Figura 3.2 - Algoritmo gerado para acesso ao banco de dados BioterC. Criado para acesso aos dados a partir do SQL.	13
Figura 3.3 - Importação de tabela contendo dados referentes aos registros de nascimento da colônia presentes no BioterC em arquivo de formato .CSV. Tabela contendo principais informações de nascimento de animais.	15
Figura 3.4 - Algoritmo para gerar gráficos do tipo barras, histograma e boxplot a partir da soma total do número de camundongos e por linhagem. Visualização da distribuição dos dados de nascimento.	17
Figura 3.5 - Algoritmo para identificação dos valores de produção animal.	18
Figura 3.6 - Algoritmo para retirar outliers identificados no conjunto de dados entre registros de nascimento. Valores que estavam abaixo ou acima do limite máximo e mínimo encontrado entre os dados foram filtrados a partir do cálculo dos quartis.	20
Figura 3.7 - Determinação de DataFrame para construção da série temporal a partir dos dados de nascimento de animais. Importação de bibliotecas para decomposição da série temporal.	21
Figura 3.8 – Algoritmo de decomposição da série temporal a partir dos dados mensais de nascimento de camundongos para todo conjunto de dados. Grupos Isogênico e Heterogênico. Decomposição da tendência, sazonalidade e ruído.	23
Figura 3.9 - Algoritmo com valores medianos de nascimento para os grupos Isogênico e Heterogênico. Apresentado em gráfico de barras.	24
Figura 3. 10 - Algoritmo para calcular o intervalo entre partos e do nascimento ao desmame para todas as linhagens. Apresentado em gráfico de barras.	26
Figura 3. 11 - Algoritmo para buscar nos dados, quais as linhagens dos grupos isogênico e heterogênico mais solicitadas entre instituições internas e externas. Gráfico apresentando as três linhagens mais solicitadas entre os usuários e instituições externas que solicitaram com mais frequência.	29
Figura 3.12- Algoritmo para gerar gráficos apresentando solicitação de camundongos entre usuários externos. Os dados foram agrupados e plotados em um gráfico de barras.	30
Figura 3. 13 - Algoritmo para processamento e mineração dos motivos de descarte animal a partir dos registros de nascimento. Dados apresentados em gráficos e tabelas.	33
Figura 4.1 - Distribuição do número de animais a partir dos registros de nascimento (A) Representação por histograma da distribuição de nascimentos de camundongos por evento de nascimento. (B) Soma total de camundongos registrados no BioterC separados por cada linhagem produzida.	35
Figura 4.2 - Identificação de valores discrepantes (outliers) no conjunto de dados de registros de nascimentos. Gráfico representando a variação de dados observados para o número de animais produzidos. Cada ponto representa uma entrada única no banco de dados de registro de nascimentos.	38
Figura 4.3 - Distribuição do número de animais a partir dos registros de nascimento após a retirada de outliers. (A) Histograma após a retirada dos outliers apresenta uma curva simétrica na distribuição dos dados. (B) Após a retirada dos valores discrepantes, as linhagens C57BL/6 e BALB/c foram mais frequentemente produzidas, com uma diferença de 101 camundongos nascidos.	41

Figura 4.4 - Decomposição em série para os nascimentos de animais somando os meses de cada ano para todo o conjunto de dados. Representação gráfica do total de nascimentos na série temporal (vermelho) apresentando aumento gradativo, decomposição dos componentes de tendência (verde) acompanhando esse crescimento, sazonalidade (azul) e ruído (cinza).	44
Figura 4.5 - Decomposição em série para os nascimentos de animais somando os meses de cada ano para o grupo Isogênico. Representação da decomposição em série entre os meses dentro do período de produção do grupo isogênico do biotério. O número de nascimentos (vermelho), tendência (verde) acompanhando esse crescimento, sazonalidade (azul), ruído (cinza).	45
Figura 4.6 - Decomposição em série para os nascimentos de animais somando os meses de cada ano para o grupo Heterogênico. Representação da decomposição em série entre os meses dentro do período de produção do grupo Heterogênico do biotério. O número de animais (vermelho) apresentando crescimento até o ano de 2018, tendência (verde) acompanhando o crescimento até 2018 e posteriormente a queda em 2019, sazonalidade (azul), aleatoriedade (cinza).	47
Figura 4.7 - Mediana de nascimentos segundo linhagem. (A) O grupo de linhagens isogênicas possui um número menor de nascimentos em comparação ao grupo de linhagens heterogênicas. (B) A Linhagem Swiss (heterogênica) apresenta valores médios e medianos mais elevados do que as isogênicas.	48
Figura 4.8 - Mediana do Intervalo entre partos para todas as linhagens. A maioria das linhagens do grupo isogênico apresentam valores superiores à média e mediana entre este período. As linhagens do grupo heterogênico requerem aproximadamente 22 a 30 dias para o nascimento.	49
Figura 4.9 - Mediana do período entre o nascimento ao desmame para todas as linhagens. A maioria das linhagens se apresentam período de nascimento e o desmame em 19 dias.	50
Figura 4.10 - Linhagens mais solicitados ao LACEA entre usuários internos e externos. (A) Linhagens frequentemente solicitadas entre os pesquisadores internos do Instituto Carlos Chagas, sendo Balb/c a linhagem mais utilizada entre diferentes laboratórios. (B) Linhagens frequentemente solicitadas entre os usuários de instituições externas, são C57BL/6 e Balb/c.	51
Figura 4.11- Instituições externas que mais solicitaram animais ao LACEA no período de cinco anos. Entre as instituições externas que solicitam camundongos pelo LACEA, a mais frequente é a UFPR e em seguida UEL.	53
Figura 4.12- Principais motivos de descarte animal registrados pelo LACEA. (A) Total de animais que foram descartados para cada motivo de descarte. (B) Total de animais descartados por motivos de descarte para o grupo isogênico. (C) Total de animais descartados por motivos de descarte para o grupo heterogênico.	57
Figura 4.13- Principais motivos de descarte da linhagem C57BL/6. (A) Representação da soma de todos os motivos de descarte registrados para a linhagem C67BL/6 (B) Soma do descarte de camundongos machos C57BL/6. (C) Soma do descarte de camundongos C57BL/6 fêmeas.	61

LISTA DE TABELAS

Tabela 3.1 - Informações das colunas utilizadas para análise exploratória dos dados, a partir da tabela est_est_log.....	14
Tabela 3.2 - Informações das colunas utilizadas para análise dos dados a partir das solicitações, armazenadas na tabela pedido.....	27
Tabela 4.1 - Distribuição do número de animais considerando como unidade observacional o registro de produção anual e o número de nascimentos (total e para cada linhagem).	36
Tabela 4.2 - Identificação de <i>outliers</i> nos registros de nascimento total e para cada linhagem.	39
Tabela 4.3 - Após a identificação dos <i>outliers</i> , os valores discrepantes foram retirados e apresentados novamente a distribuição dos valores de nascimento.	40
Tabela 4.4 - Distribuição do número de animais considerando como unidade observacional o registro de produção anual e o número de nascimentos (total e para cada linhagem sem a presença de <i>outliers</i>).	42
Tabela 4.5 - Número de nascimentos mensais distribuídos anualmente para todo o conjunto de dados.	44
Tabela 4.6 - Número de nascimentos mensais distribuídos anualmente para o grupo Isogênico.	46
Tabela 4.7 - Número de nascimentos mensais distribuídos anualmente para o grupo Heterogênico.	47
Tabela 4.8 - Número de entradas e de camundongos solicitados ao LACEA distribuídos anualmente.	50
Tabela 4.9 - Número de entradas e soma de animais de cada linhagem solicitada entre os usuários internos e externos.	52
Tabela 4.10 - Número de entradas e a soma de animais solicitados ao LACEA por instituições externas.	53
Tabela 4.11 - Número de entradas e de camundongos descartados distribuídos anualmente.	54
Tabela 4.12 - Número de entradas e soma de animais segundo os principais motivos de descarte animal para todo conjunto de dados.	55
Tabela 4.13 - Número de entradas e soma de animais segundo os principais motivos de descarte animal para o grupo Isogênico.....	58
Tabela 4.14 - Número de entradas e soma de animais segundo os principais motivos de descarte animal para o grupo Heterogênico.....	59
Tabela 4.15 - Número de entradas e soma de animais segundo os principais motivos de descarte animal para a linhagem C57BL/6.....	62
Tabela 4.16 - Número de entradas e soma de animais segundo os principais motivos de descarte animal para machos e fêmeas da linhagem C57BL/6.....	63
Tabela 4.17 - Registros na sessão de descarte do BioterC que foram representados como “Outros” por não serem motivos de descarte animal.	64

LISTA DE QUADROS

Quadro 8.1 – Termologias utilizadas em linguagens de programação	80
Quadro 8.2- Biblioteca Pandas de Python	81
Quadro 8.3 - Biblioteca Numpy em Python.....	81
Quadro 8.4 - Biblioteca Statsmodels em Python	81
Quadro 8.5 - Biblioteca Seaborn e Matplotlib em Python.....	81

LISTA DE ABREVIATURAS E SIGLAS

CCAC	<i>Canadian Council on Animal Care</i>
CEUA	Comissão de Ética no Uso de Animais.
CITS	Centro Internacional de Tecnologia de Software
CONCEA	Conselho Nacional de Controle de Experimentação Animal
FELASA	<i>Federation of European Laboratory Animal Science Associations</i>
IDE	<i>Integrated Development Environment</i>
ILAR	<i>Institute for Laboratory Animal Resources</i>
LACEA	Laboratório de Criação e Experimentação Animal.
IEP	Intervalo Entre Partos
SPF	<i>Specific Pathogen Free</i>
SQL	<i>Structured Query Language</i>

Laboratórios internos Instituto Carlos Chagas:

BIOMOLTRIP	Laboratório Biologia molecular de Tripanossomatídeos
LABCET	Laboratório de Biologia Básica de Células Tronco
LABGEN	Laboratório de Genômica
LABREG	Laboratório de Regulação e Expressão Gênica
LBC	Laboratório de Biologia Celular
LPEP	Laboratório de Biologia Estrutural e Engenharia de Proteínas
VIROMOL	Laboratório de Virologia

Instituições Externas:

UFSC	Universidade Federal de Santa Catarina
PUC PR	Pontifícia Universidade Católica do Paraná
UFRJ	Universidade Federal do Rio de Janeiro
UFFS	Universidade da Fronteira Sul
UFPR	Universidade Federal do Paraná
UFRJ	Universidade Federal do Rio de Janeiro
UFSC	Universidade Federal de Santa Catarina
UFSM	Universidade Federal de Santa Maria
UFSM	Universidade Federal de Santa Maria
IBMP	Instituto de Biologia Molecular do Paraná
USP	Universidade de São Paulo



Ministério da Saúde

FIOCRUZ - PARANÁ
Instituto Carlos Chagas

Ata da Sessão Pública de exame de dissertação para obtenção do grau de Mestre em
Biociências e Biotecnologia.

Aos 26 dias do mês de fevereiro de dois mil e vinte e um, às 14:00 horas, **através de teleconferência pelo link** <https://us02web.zoom.us/j/5953384663?pwd=aVFKSU52ZnNRbGVQOHZzNkxKNzZGUT09>, reuniu-se a Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Biociências e Biotecnologia, composta pelos Professores: Dr. Leonardo Foti, Dra. Hellen Geremias dos Santos e Dr. Marcos Antonio Maioli, com a finalidade de julgar a dissertação da candidata **Eloiza Kauanna Gonçalves Dias Ferreira**, intitulada: **“Uso de ferramentas de Ciência de Dados para avaliação dos dados zootécnicos de produção de linhagens de camundongos: uma proposta de melhorias ao gerenciamento do plantel”**, sob a orientação de **Dr. Guilherme Ferreira Silveira**, para obtenção do grau de **Mestre** em Biociências e Biotecnologia. A candidata teve até 45 (quarenta e cinco) minutos para a apresentação, e cada examinador teve um tempo máximo de arguição de 30 (trinta) minutos, seguido de 30 (trinta) minutos para resposta do(a) candidato(a) ou de 60 (sessenta) minutos quando houve diálogo na arguição. O desenvolvimento dos trabalhos seguiu o roteiro de sessão de defesa, estabelecido pela Coordenação do Programa, com abertura, condução e encerramento da sessão solene de defesa feito pelo Presidente **Dr. Leonardo Foti**. Após haver analisado o referido trabalho e arguido o(a) candidato(a), os membros da banca examinadora deliberaram pela:

Aprovação da dissertação por unanimidade

Aprovação somente após satisfazer as exigências de modificações no prazo fixado pela banca (não superior a noventa dias).

Reprovação da dissertação

Na forma regulamentar foi lavrada a presente ata que é abaixo assinada pelos membros da banca, na ordem acima determinada, e pela candidata.

Curitiba, 26 de fevereiro de 2021

Dr. Leonardo Foti – ICC/FIOCRUZ

Dra. Hellen Geremias dos Santos – ICC/FIOCRUZ

+55(41)3316-3230

Rua Prof. Algacyr Munhoz Mader, 3775 - CIC 81350-010 Curitiba/PR, Brasil

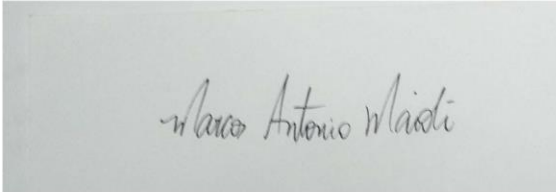




Ministério da Saúde

FIOCRUZ - PARANÁ
Instituto Carlos Chagas

Dr. Marcos Antonio Maioli – CEDEME/UNIFESP



Candidato(a): Eloiza Kauanna Gonçalves Dias Ferreira

+55(41)3316-3230

Rua Prof. Algacyr Munhoz Mader, 3775 - CIC 81350-010 Curitiba/PR, Brasil



Sumário

1. Introdução.....	1
1.1. Histórico de Criação e Experimentação Animal	1
1.2. Laboratório de Criação e Experimentação Animal	4
1.3. Gerenciamento animal através de <i>softwares</i>	6
1.4. BioterC	7
1.5. Ciência de dados e modelagem computacional.....	8
2. Objetivo Geral.....	10
2.1. Objetivos específicos	10
3. Materiais e Métodos	10
3.1. Processamento dos dados obtidos pelo software BioterC	10
3.2. Organização do banco de dados do BioterC em SQL (do inglês, <i>Structured Query Language</i>)	11
3.3. Linguagem de programação Python	12
3.3.1. Bibliotecas em Python utilizadas para análise exploratória dos dados. 12	
3.3.2. Ambiente de análise, repositório e Ambiente de Desenvolvimento Integrado (do inglês <i>Integrated Development Environment - IDE</i>).....	12
3.4. Lógica construída para os algoritmos	13
3.4.1. Acesso ao banco de dados	13
3.4.2. Tabela utilizada para análise	13
3.4.3. Soma total do número de camundongos produzidos e por cada linhagem	15
3.4.4. Decomposição da Série Temporal para todo o conjunto de dados de nascimentos e para os grupos Isogênico e Heterogênico	20
3.4.5. Mediana da produção de camundongos para os grupos Isogênicos e Heterogênicos.....	23
3.4.6. Mediana do intervalo entre partos (IEP) e do nascimento ao desmame	24
3.4.7. Linhagens solicitadas com mais frequência por usuários internos e externos.....	26
3.4.8. Instituições externas que mais solicitaram camundongos	29
3.4.9. Motivos de descarte dos camundongos em todo o conjunto de dados, para os grupos Isogênicos e Heterogênicos e para a linhagem C57BL/6	30
4. Resultados.....	34
4.1. Conjunto de dados do BioterC apresenta valores discrepantes para nascimento de animais	34
4.2. A série temporal indica maior produção de camundongos em 2018.....	43

4.3.	Produção do grupo isogênicas apresentam maior produção em 2018.	45
4.4.	Produção de linhagens heterogênicas apresentam maior produção no ano de 2016.....	46
4.5.	O grupo Isogênico apresenta mediana de nascimentos inferior aos heterogênicos.	48
4.6.	A mediana do intervalo entre partos é em torno de 35 dias.....	49
4.7.	A mediana entre o nascimento e o desmame dos camundongos é de 19 dias.	49
4.8.	As linhagens que são mais solicitadas entre usuários internos e externos do LACEA são C57BL/6 e Balb/c.....	50
4.9.	O principal motivo de descarte de camundongos é por Descarte zootécnico.	54
4.10.	As análises dos dados do BioterC direcionaram modificações gerenciais no LACEA.	65
5.	Discussão	68
6.	Conclusões.....	73
7.	Referências	74
8.	Apêndice.....	80
8.1.	Quadros apresentando funções e termos computacionais	80
8.2.	Métodos utilizados para Análise Exploratória de Dados	81
9.	Artigo	82

1. Introdução

1.1. Histórico de Criação e Experimentação Animal

A pesquisa científica dentro da área biológica contribui de maneira irrefutável para a saúde e bem-estar dos seres humanos e animais. Há séculos, animais vêm sendo utilizados como modelo vivo em estudos. Inicialmente, esses modelos foram usados em trabalhos de anatomia e fisiologia que envolviam a dissecação para análise das funções corporais (Simmons, 2004). Durante o século XIX, vários zoólogos europeus criaram camundongos para investigar caracteres genéticos e tentaram interpretar os resultados da *lei de Galton* sobre a herança ancestral. As informações valiosas adquiridas não puderam ser corretamente interpretadas até a redescoberta do trabalho de Mendel, em 1900 (Keeler, 1931).

Diferentes países têm buscado a regulamentação para o controle do uso de animais na pesquisa, garantindo assim seu uso responsável e ético.

Os princípios dos 3Rs (*Replacement, Reduction e Refinement*) existem há mais de 50 anos, e fornecem uma estrutura para a realização de pesquisas com animais. Desde então, eles foram inclusos à legislação e aos regulamentos nacionais e internacionais sobre o uso de animais em procedimentos científicos, bem como nas políticas de organizações que financiam ou conduzem pesquisas com animais (NC3Rs, 2021).

Com base nessas premissas, em 1952, foi criado no Estados Unidos o Conselho Nacional de Pesquisa com o *Institute for Laboratory Animal Resources* (ILAR), com o intuito de coordenar informações sobre o uso de animais de laboratório, incluindo departamentos e agências que usam animais para pesquisa. O objetivo principal do ILAR é avaliar e relatar o uso científico, tecnológico e ético de animais de laboratório e o uso de métodos alternativos, em ambientes não alimentares, tais como pesquisa, educação e produção de produtos farmacêuticos (NRC, 1988).

Em 1968, foi fundado o *Canadian Council on Animal Care* (CCAC), responsável por definir, manter e supervisionar a implementação de altos padrões de ética, bem-estar e cuidado animal na ciência em todo o Canadá (CCAC, 2019).

Em 1978, foi fundada a *Federation of European Laboratory Animal Science Associations* (FELASA), que tem como objetivo representar interesses comuns em todos os aspectos para animais de laboratório na Europa, defendendo uma conduta

científica responsável com os animais nas ciências da vida, com particular ênfase na garantia do bem-estar animal (Prins, 2018).

No Brasil, foi inaugurado em 1904 o Instituto Soroterápico do Rio de Janeiro, atual Fundação Oswaldo Cruz, que abrigava o primeiro biotério de produção animal do país. Transformações nas condutas de pesquisa e avanços tecnológicos observados ao longo de 116 anos levaram à necessidade de melhorias nas instalações devido à crescente utilização de animais para o ensino e a pesquisa (Cruz, 2003).

A promulgação da Lei 11794/2008 (Lei Arouca), posteriormente regulamentada pelo Decreto 6899/2009, promoveu uma busca por adequação levando a mudanças no cenário da experimentação animal no Brasil (Mazzarotto; Silveira, 2013). A legislação indica a necessidade de melhorias no gerenciamento dos biotérios e a padronização das condições de manutenção e de experimentação, além da padronização genética e sanitária. Em 2008, foi criado o Conselho Nacional de Controle de Experimentação Animal (CONCEA) órgão de caráter normativo, consultivo, deliberativo e recursal, que norteiam a devida aplicação da Lei Arouca. O CONCEA prevê ainda a criação e funcionamento dos Comitês de Ética no Uso de Animais (CEUA), que estão presentes nas instituições de pesquisa científica que utilizam animais de laboratório (Dalben; Emmel, 2013).

No tocante à padronização dos biotérios, destaca-se que os animais devem ser criados e mantidos em condições físicas e ambientais favoráveis e de biossegurança.

A infraestrutura deve dispor de programas de gerenciamento, condições físicas e ambientais, garantindo cuidados que beneficiem o desenvolvimento e a reprodução dos animais, mantendo o bem-estar, além de minimizar variações que podem interferir nos resultados dos ensaios (NIH, 2002). Entre os padrões de infraestrutura, existe o microambiente, que se refere ao espaço físico do animal, ou seja, a caixa ou gaiola, com parâmetros próprios para temperatura, umidade relativa, composição de gases e partículas do ar. Já o macroambiente, se refere ao ambiente físico secundário (em torno do microambiente), representado pela sala de criação, o estábulo ou qualquer outro hábitat externo (NRC, 1976).

O nível de biossegurança é determinado através do planejamento estrutural das instalações, de níveis de contenção física, além da utilização apropriada de

equipamentos de proteção individual e/ou coletivos e emprego dos procedimentos de boas práticas laboratoriais (CDC, 1979).

Em biotérios que realizam experimentação, é adequado que a edificação esteja próxima ao laboratório de pesquisa, com barreiras sanitárias de proteção, tanto para o bioterista quanto para o meio ambiente. Em relação aos biotérios de criação, é conveniente que sejam instalados em áreas distantes de centros urbanos, isolando os animais de agentes biológicos que possam infectá-los ou contaminá-los, garantindo assim, o cumprimento das normas sanitárias adequadas de criação e manutenção (Fiocruz, 1998; Cardoso, 2001).

Ainda com relação à produção de animais em biotérios de criação, devem ser consideradas as características genéticas, que são de grande importância no que se refere à seleção e ao manejo dos animais utilizados em pesquisas na área da saúde (NRC, 2003). Estes protocolos de padronização são conhecidos como padrão genético e sanitário.

Conforme a caracterização genética dos animais, existem dois tipos de linhagens: isogênicas (*inbred*) e heterogênicas (*outbred*) (Jonas, 1976; Festing, 1993). Um aspecto importante relacionado ao padrão genético é a confiança e reprodutibilidade de dados obtidos com linhagens dos animais utilizados na experimentação. Experimentos realizados com animais contaminados (isto é, de herança genética indeterminada), podem comprometer a exatidão dos resultados, dificultando sua interpretação ou fornecendo resultados falsos positivos, tornando esse animal inadequado como modelo experimental (Passos, 2003).

O padrão sanitário é outro controle importante para a criação e uso de animais de laboratório de qualidade. Recentemente, os critérios na pesquisa científica têm exigido animais com padrão sanitário definido, de modo a aumentar a confiança da obtenção dos dados e, conseqüente, reprodutibilidade da pesquisa. Para se atingir este padrão, primariamente devem ser realizados investimentos na melhoria da estrutura dos biotérios, principalmente nos de criação. Com a padronização sanitária, tem-se conseguido diminuir o número de animais usados em experimentação, sobretudo na diminuição das variações nos grupos de animais-teste, contribuindo tanto para o bem-estar dos animais quanto dos profissionais envolvidos neste processo, uma vez que diminuem os riscos de zoonoses (FELASA, 1999). Neste contexto, Majerowicz (2005) categorizou os seguintes padrões sanitários:

- “Convencionais”: são criados em sistemas de gaiolas abertas com fluxo livre de pessoas e materiais. Não apresenta qualquer tipo de barreira sanitária que possa impedir a introdução de agentes externos, tornando-os susceptíveis a contaminações e infecções;
- “Livres de Microrganismos Patogênicos Específicos” (*specific pathogen free - SPF*): são criados em biotérios com eficientes barreiras sanitárias ou são mantidos em acomodações que evitam contato com potenciais agentes patogênicos, tornando sua microbiota controlada. Esse padrão de saúde irá depender da lista individual de exclusão de micro-organismos de cada biotério;
- “Axênicos ou *Germfree*”: são criados e mantidos em isoladores que os mantém livres de quaisquer tipos de microrganismos ou outras formas de vida associadas;
- “Mantidos em Barreiras”: são criados como tendo a microbiota definida, porém, posteriormente são retirados dos isoladores e alocados em um biotério com barreiras sanitárias definidas, para que sejam monitorados quanto à presença de micro-organismos inoculados ou acidentalmente adquiridos;
- “Monitorados”: são criados e alojados em sistema de barreira de baixa segurança e através de monitoramento periódico, apresentam-se livres da maioria dos patógenos.

A situação brasileira, quanto à padronização dos processos (genéticos, sanitários, de biossegurança, etc.), não é satisfatória, o que pode ser observado pela escassa literatura especializada, direcionada à educação e treinamento dos profissionais envolvidos (Cruz, 2003). Nos raros biotérios considerados padronizados, algumas atividades têm sido realizadas por técnicos, quando, de fato, seriam encargos essencialmente gerenciais (Cruz, 2003). Assim, fica evidente que no Brasil é recente a preocupação com a padronização dos biotérios. A atenção dada e a conscientização da importância de biotérios adaptados aos padrões de qualidade exigidos por guias internacionais, ainda são primários (Duarte, 2003).

1.2. Laboratório de Criação e Experimentação Animal

O Laboratório de Criação e Experimentação Animal (LACEA) do Instituto Carlos Chagas (ICC), foi inaugurado em 2014, contando com uma equipe de oito

colaboradores capacitados em bioterismo. A área possui todos os padrões necessários para classificar o LACEA como sendo um biotério de barreiras.

Um exemplo desta característica é a presença de espaços bem definidos no local, tais como área de criação de animais, área de experimentação, área de lavagem e salas de apoio e de circulação (sala quarentena, salas de estoque, sala de freezers, banheiros, secretaria, corredores, entre outros). O LACEA possui ainda dois pisos técnicos onde estão alojados o parque de máquinas de climatização e os dutos de ar-condicionado centrais (Diretrizes Biotério LACEA, 2017).

O foco do LACEA, no cenário da experimentação animal no Brasil, é fomentar a pesquisa biomédica brasileira por meio do fornecimento de animais de laboratório com qualidade, distribuição de insumos de alto valor biotecnológico agregado e atuação na formação de recursos humanos em diferentes áreas do laboratório.

O LACEA atende todos os laboratórios do ICC, sendo que, destes, seis com maior frequência (LABCET, LPEP, LABREG, LABGEN, VIROMOL e BIOMOLTRIP), um Instituto Parceiro (IBMP) e tem atendido de forma esporádica instituições externas (Biovet, Immunova, UFSC, UEL, UFRJ, USP, UFPR, etc.) (Diretrizes LACEA, 2017).

O fornecimento de animais para laboratórios internos e externos ocorre depois que o solicitante apresenta o documento de Comissão de Ética no Uso de Animais (CEUA), contendo as informações, considerando-se: a espécie e linhagem animal, sexo, quantidade e vigência do certificado. Os animais são fornecidos em conformidade com a disponibilidade dos biotérios e dentro do limite de validade do certificado da CEUA.

Os animais do LACEA são considerados padrão SPF e, para manter a padronização sanitária e genética, o biotério segue as recomendações da FELASA “Monitoramento da saúde de colônias de camundongos, ratos, hamsters, porquinhos-da-índia e coelhos em unidades de reprodução e experimentais”, publicado em 2014. Para a gestão de qualidade, seguindo as normas – ABNT NRB ISO 9001:2015. O LACEA é dotado dos aspectos construtivos necessários para atender às normas de biossegurança e boas práticas para a produção de animais de laboratório (Diretrizes Biotério LACEA, 2017).

O LACEA produz colônias de animais de diferentes linhagens que podem ser classificados nos grupos isogênicos e heterogênicos, de acordo com sua genética, definida através de práticas reprodutivas sistemáticas.

Camundongos do grupo heterogênico são animais que possuem constituição genética variada através de acasalamento randômico (aleatórios), evitando que animais em acasalamento sejam parentes próximos. Estes animais são mais robustos, apresentam maior fertilidade, com maior número de filhotes por ninhada, sendo exemplo desse grupo a linhagem Swiss (Santos, 2002).

Entre as linhagens isogênicas, para garantir pureza da linhagem e homogeneidade, é preciso realizar acasalamentos de, no mínimo, 20 gerações subsequentes, a partir de um único casal, resultando em um coeficiente de endogamia de 98,6% (Santos, 2002). Esse grupo apresenta menor fertilidade e, portanto, em decorrência disso, uma menor prole. As linhagens desse grupo permitem a realização de experimentos eliminando a variabilidade de origem genética, assegurando assim a utilização de menos animais para que testes estatísticos apresentem poder adequado (Casellas, 2010).

1.3. Gerenciamento animal através de *softwares*

Com o intuito de facilitar melhorias e padronizações, o uso de *softwares* específicos para o gerenciamento de biotérios tornou-se uma opção atrativa para o armazenamento, manutenção e recuperação dos dados de gerenciamento da criação das colônias. Estes programas proporcionam melhor visualização da alocação dos recursos financeiros e facilitam tomadas de decisões gerenciais mais rápidas e seguras.

Em 2016, o ICC em parceria com a empresa Wylinka, que tem como objetivo desenvolver instituições e ecossistemas para a inovação e o empreendedorismo, realizaram um levantamento de *softwares* capazes de gerenciar a produção animal de biotérios:

- MosaicVivarium, oferece uma solução abrangente para o gerenciamento de recursos de animais de laboratório como organização dos animais, rastreamento de linhagens e reprodução, etc.; (Estados Unidos);
- SoftMouse, apresenta funcionalidades como estatísticas das colônias, mapeamento da linhagem, planejamento de experimentos.; (Estados Unidos e Canadá)

- AniBio, projetado para um gerenciamento eficiente do biotério e dos animais, registro de criação de animais, genotipagem, rastreabilidade do histórico do animal, etc; (Espanha)
- PyRat, apresenta funcionalidades como gerenciamento de reprodução, relatórios estatísticos, dados de genotipagem etc.; (Alemanha).

A utilização de softwares especializados nos biotérios brasileiros é pontual. Dado o fato de a maioria das soluções de mercado serem comercializadas como produto, os valores praticados por estas desenvolvedoras não se apresenta atrativo. Alguns dos softwares acima citados são vendidos na faixa de R\$ 150.000,00. Adicionalmente, não existem soluções nacionais para a demanda, o que também é um dificultador na elaboração de um valor competitivo para o mercado brasileiro. Assim, somando-se as informações acima, o desenvolvimento de uma solução própria torna-se necessária (Wylinka, 2016).

1.4. BioterC

Com o objetivo de melhorar a capacidade do processo gerencial do LACEA, foi desenvolvido e implementado o *software* BioterC (INPI - BR512014001421-5), um sistema de gerenciamento flexível, com diferentes funcionalidades para criação de animais. O BioterC armazena, desde sua implementação em setembro de 2014, todos os registros de entradas de movimentação no LACEA até junho de 2019. No sistema constam informações sobre as colônias criadas de todas as linhagens, bem como registros de procedimentos rotineiros do LACEA como acasalamento, registro de nascimento, período pós desmame, identificado nesse software como sexagem (separação de camundongos machos e fêmeas), fornecimento e descartes zootécnicos. Adicionalmente, também estão inseridas informações do controle do LACEA, como estoque de animais e insumos para criação e manipulação dos animais, entre outras funções.

O BioterC facilitou a transmissão padronizada das informações sobre os regulamentos internos do biotério, além de ter gerado acréscimo na capacidade de previsibilidade do fornecimento e produção dos animais. Ademais, o uso desse *software* reduziu o tempo necessário para a produção de relatórios para os órgãos de controle da atividade de criação e experimentação animal. Em linhas gerais, o BioterC

tem proporcionado segurança para a condução das rotinas de criação e de experimentação animal, mostrando transparência e agilidade para a extração de todos os dados que permeiam o criatório animal (Mazzarotto; Silveira, 2013)

Em junho de 2019, o BioterC foi substituído por uma nova versão, o BioterCS. Nele, as funcionalidades foram atualizadas e são capazes de se modificar a partir da necessidade de cada biotério.

1.5. Ciência de dados e modelagem computacional

Esta área é baseada em princípios computacionais, matemáticos/estatísticos e de conhecimento da área de pesquisa que geram as informações e tem como objetivo identificar padrões contidos nos próprios dados e a aplicação de métodos quantitativos e qualitativos para previsibilidade de problemas e resultados. Capaz de transformar grande conjunto de dados brutos em *insights*, auxiliando em tomadas de decisões para atingir melhores resultados (Figura 1.1).



Figura 1.1 - Diagrama de conhecimentos que agregados dão origem à área de Ciência de Dados. Fonte: Medium, 2018.

Conceitualmente, a Ciência de Dados pode ser dividida em diferentes áreas e etapas reconhecidas por conhecimentos e ações específicos. Dentre elas, o processo de:

- i. Entendimento do problema: que consiste em fazer perguntas relevantes para serem respondidas através do conjunto de dados;
- ii. Coleta dos dados: refere-se a obter o conjunto de dados que será analisado;
- iii. Processamento dos dados: baseia-se na estruturação dos dados, que muitas vezes não estão organizados;
- iv. Análise exploratória de dados (AED): busca entender o conjunto de dados, utilizando conceitos como medidas resumo de tendência central e de variabilidade. A AED contribui para a formulação de hipóteses que podem auxiliar na tomada de decisões;
- v. Comunicação dos Resultados: uma vez obtidas as informações via AED, estas podem ser apresentadas utilizando-se gráficos e dashboards;
- vi. Feedback: tomada de decisões a partir dos resultados alcançados.

A extração do conhecimento agregado em padrões dos dados por meio de tecnologias que incorporam princípios de inteligência artificial é conhecida como modelagem computacional. Existem centenas de diferentes algoritmos de mineração de dados e uma grande quantidade de detalhes para os métodos do campo (Provost, 2013). Dentre esses, a análise estatística de Séries Temporais em que os dados são observados em diferentes instantes do tempo, definidas como “sequências de dados quantitativos relativos a momentos específicos e estudados segundo sua distribuição no tempo” (Wiener, 1966), foram utilizadas na exploração dos dados do BioterC.

O presente trabalho se justifica pela busca constante de resultados experimentais relevantes e reprodutivos na criação animal. Pela melhoria do bem-estar animal, a otimização dos procedimentos da criação, visando contribuir com a utilização consciente e na implementação de medidas capazes de reduzir o descarte zootécnico de animais. Adicionalmente, o presente trabalho tem como justificativa secundária colaborar na criação e fortalecimento de um grupo de Ciência de Dados no Instituto Carlos Chagas – Fiocruz/PR.

Ferramentas de Ciência de Dados permitem identificar padrões na produção e utilização de camundongos.

2. Objetivo Geral

Identificar e propor mudanças nas rotinas de manejo da colônia através das análises do conjunto de dados obtidos no *software* de gerenciamento BioterC, desenvolvido no LACEA, utilizando ferramentas de Ciência de Dados.

2.1. Objetivos específicos

- Identificar padrões que sejam capazes de caracterizar a criação animal no biotério do ICC;
- Analisar séries temporais de produção de animais dentro do período em que os dados foram registrados;
- Identificar entre os grupos genéticos isogênicos e heterogênicos, quais as linhagens de camundongos mais solicitadas ao LACEA, bem como instituições externas responsáveis por essas solicitações;
- Descrever os principais motivos de descarte animal através dos dados registrados no BioterC.

3. Materiais e Métodos

3.1. Processamento dos dados obtidos pelo software BioterC

A Figura 3.1, a seguir, apresenta o fluxo de dados e como foi realizada cada etapa das análises, que foram baseadas no “Ciclo de Vida da Ciência de Dados” citado anteriormente. O primeiro passo foi a aquisição dos dados armazenados em formato SQL (do inglês, *Structured Query Language*, vide tópico 3.2). Para as análises foi utilizado linguagem de programação *Python* (vide tópico 3.3), e a escrita desse código foi realizada no programa Anaconda sendo utilizada a IDE *Jupyter Lab 6.2.0* (vide tópico 3.4). Foi necessário realizar a estruturação e organização desse conjunto de dados para que fosse possível construir uma lógica e determinar algoritmos para gerar os resultados. A última etapa para conclusão desse ciclo, foi a tomada de decisões, verificando se nossos resultados estavam de acordo com o que é observado na rotina do biotério e na literatura.

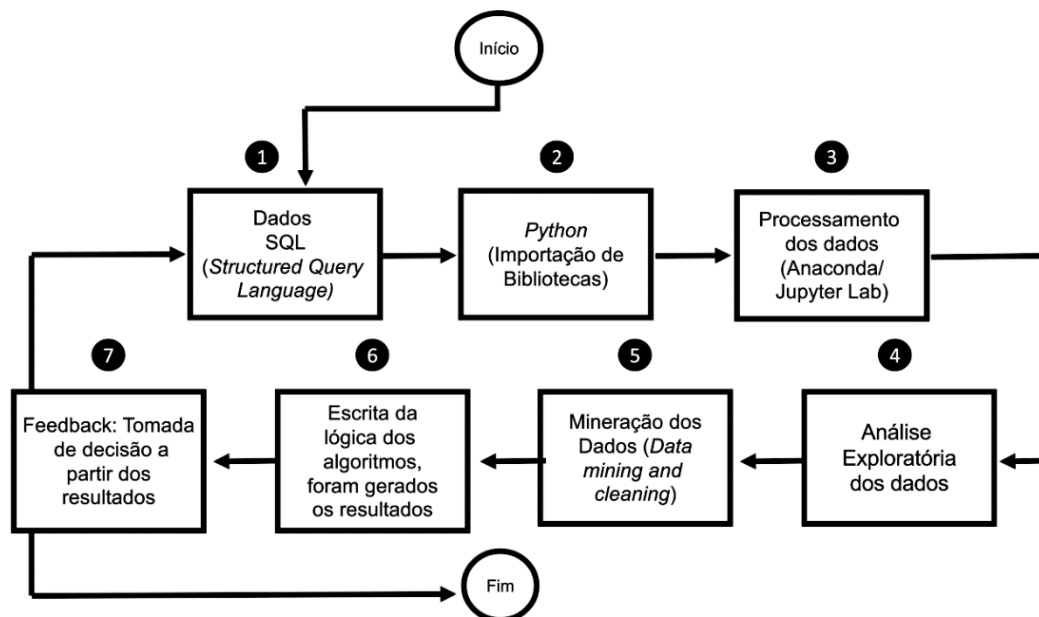


Figura 3.1 - Fluxograma do processamento de dados do BioterC.

3.2. Organização do banco de dados do BioterC em SQL (do inglês, *Structured Query Language*)

Os dados do sistema BioterC foram acessados diretamente nos servidores da empresa que atualmente tem a cessão de exploração comercial do BioterC, Centro Internacional de Tecnologia de Software (CITS). Esses dados originalmente estão em formato SQL, um sistema gerenciador de banco de dados relacional, que tem como principal função permitir o armazenamento e a recuperação estruturada de informações. Estas podem ser solicitadas por outras aplicações de *software*, seja no mesmo servidor ou em outra instância através de uma rede de internet. Ao solicitar o banco, tivemos acesso a 12 tabelas referentes ao gerenciamento da colônia do LACEA. Cada tabela apresenta um total de registros, que serão apresentadas nos próximos tópicos.

3.3. Linguagem de programação Python

Buscando realizar análise dos dados obtidos do *software* BioterC, foi utilizada a linguagem de programação *Python*, utilizando-se de conceitos de AED. *Python* trata-se de uma linguagem de sintaxe comparativamente simples e de ampla utilização em análise de dados, frente a outras linguagens, além de possuir diferentes bibliotecas como ferramentas para análises e estruturas de dados de alta performance, como *Pandas*, *Numpy*, *Matplotlib* e *Seaborn*.

3.3.1. Bibliotecas em Python utilizadas para análise exploratória dos dados

- *Pandas* é uma biblioteca que fornece ferramentas de análise e estruturas de dados de alta performance, chamados *DataFrames*, sendo a principal e mais completa biblioteca para Ciência de Dados contida na linguagem *Python*.
- *Numpy* é uma biblioteca usada principalmente para realizar cálculos em *arrays* (estruturas de dados semelhantes aos *DataFrames*) multidimensionais. Essa biblioteca fornece um conjunto de funções e operações que auxilia na execução de cálculos numéricos e estatísticos.
- *Matplotlib* é utilizada para gerar gráficos com alta qualidade a partir de comandos simples.
- *Seaborn*, foi construído utilizando as bases da biblioteca *Matplotlib* e contribui com a visualização dos dados gráficos, otimizando a exposição da informação.
- *Statsmodels* é um modulo utilizado em python que permite aos usuários explorar dados, estimar modelos estatísticos e executar testes estatísticos.

3.3.2. Ambiente de análise, repositório e Ambiente de Desenvolvimento Integrado (do inglês *Integrated Development Environment - IDE*)

O ambiente de análise foi gerado pelo programa Anaconda, uma distribuição livre e de código aberto que otimiza a utilização da linguagem de programação *Python*, e que tem como objetivo simplificar o gerenciamento de pacotes de

desenvolvimento. Os dados analisados foram armazenados em um repositório *online* de acesso privado, o *GitHub*, com um controle de versão dos estudos realizados, permitindo que usuários cadastrados na plataforma contribuam para o seu desenvolvimento. O ambiente de desenvolvimento integrado (IDE), por sua vez, é um programa que reúne características e ferramentas de apoio ao desenvolvimento de *software* otimizando o processo. Nesse caso a IDE utilizada foi a *Jupyter Lab*, que faz parte do ambiente Anaconda.

3.4. Lógica construída para os algoritmos

3.4.1. Acesso ao banco de dados

Para ter acesso ao banco de dados do BioterC, foram utilizadas as bibliotecas *Pymysql* e *Pandas*. A partir delas, foi possível acessar o servidor que continha o banco de dados do BioterC, indicando para tanto informações padrão FTP como o endereço de IP (host), Usuário/Senha e nome do banco de dados (Figura 3.2).

```
import pymysql #Importando pymysql
import pandas as pd #importando pandas
from datetime import date #importando datetime

#mantendo conexão com os dados
con = pymysql.connect(host=host, #IP
                      port=port,
                      user=user, #usuário
                      passwd=passwd, #senha
                      db=db, #banco de dados
                      )
```

Figura 3.2 - Algoritmo gerado para acesso ao banco de dados BioterC. Criado para acesso aos dados a partir do SQL.

3.4.2. Tabela utilizada para análise

O banco de dados do sistema é dividido em 12 tabelas que contém todas as informações do *software*. No desenvolvimento do BioterC, os autores direcionaram

todas as informações de histórico de movimentação da colônia à uma única tabela. Para o processamento dos dados, utilizamos inicialmente a tabela que contém informações de movimentação da colônia, denominada “est_est_log”, a qual contém 28 colunas (variáveis). Observando especificamente os registros de nascimento, foram encontradas 1.808 entradas (instâncias) de valores únicos, ou seja, não apresentando duplicidade de registros. As colunas desta tabela que foram analisadas são descritas na Tabela 3.1. Para a AED, foi utilizado a biblioteca *Pandas*, e a partir dessa tabela, foi criado um *DataFrame*.

Tabela 3.1 - Informações das colunas utilizadas para análise exploratória dos dados, a partir da tabela est_est_log.

Colunas da Tabela	Descrição do que contém em cada coluna BioterC
'data_hoje'	Data em que ocorreu o registro.
'ling'	Linhagens.
'num_ani'	Número de animais.
'm_f'	Animais machos ou fêmeas
'util'	Como os animais foram utilizados (nascimento, transferência, acasalamento etc.)
'date_acas'	Data da montagem da caixa de acasalamento.
'date_nasc_acas'	Data do registro de nascimento.
'date_sex'	Data do registro da sexagem.

1.808 entradas x 8 colunas

Fonte: BioterC,2019

A tabela “est_est_log” foi importada ao ambiente, para leitura do arquivo, utilizando funções da biblioteca *Pandas*, como *read_csv*. O método converte um arquivo CSV (*Comma-separated values*) para uma estrutura de dados no formato *DataFrame*, possibilitando a partir dele a AED (Figura 3.3).

```

nume_ani    =    pd.read_csv('../tabelas_test/est_est_log_2019-06-18.csv',
index_col='id') #lendo arquivo .csv

#Criando DataFrame a partir dos filtros:

```

```

nume_ani = nume_ani[(nume_ani['num_ani'] < 200) & #Filtrando valores
inferiores a 200 para evitar de surgir valores inválidos
                (nume_ani['util'] == 'nasc') &
#Minha coluna útil cotenha apenas informações de nascimento
                (nume_ani['ling'] != 'Test')]
#Não apresente Test como uma linhagem

```

Figura 3.3 - Importação de tabela contendo dados referentes aos registros de nascimento da colônia presentes no BioterC em arquivo de formato .CSV. Tabela contendo principais informações de nascimento de animais.

3.4.3. Soma total do número de camundongos produzidos e por cada linhagem

Ao ter acesso a tabela de interesse, para entender como os dados estavam distribuídos, buscamos pela soma total de animais, e para isso, foi utilizada a coluna onde o número de nascimentos de animais foi armazenada (*num_ani*). Uma vez selecionada, utilizamos uma função de soma que nos retornou o valor total de animais. Para visualização, foi elaborado um histograma contendo a soma de animais para cada evento de nascimento. Para o gráfico de barras foi considerado esse valor total para cada linhagem que foi registrada no BioterC. Valores acima de 30 animais por nascimento sugerem que estes registros são discrepantes pois não condiziam com o que era observado na rotina do biotério e na literatura. Para confirmar a existência desses valores, foi plotado gráfico de *boxplot*, onde foi possível encontrar *outliers* por cálculo de quartis. Os valores de cada linhagem foram organizados em ordem crescente, e divididos em quartis. O primeiro quartil (Q1), é o valor que deixa 25% dos dados observados abaixo e 75% acima, enquanto o terceiro quartil (Q3), deixa 75% dos valores observados abaixo e 25% acima. O segundo quartil (Q2) representa a mediana. Para identificar os *outliers*, os valores que se apresentaram fora do limite inferior dos dados, $Q1 - 1,5 * (Q3 - Q1)$, e limite superior, $Q3 + 1,5 * (Q3 - Q1)$, considerados valores discrepantes.

Em *Python*, para calcular a soma de todos os camundongos que foram registrados como nascimento no BioterC no período analisado (contendo o primeiro registro datado em 23/09/2014 até 14/06/2019, onde estava prevista a substituição do BioterC por uma nova versão), apenas a coluna com o número de animais foi selecionada e seus valores foram somados por meio da função *sum()*.

Para visualização, esses dados somados foram plotados em um histograma e em gráfico de barras, para os quais foi necessário importar a biblioteca *Seaborn* e utilizar o método *barplot()* para gerar as barras, e o método *distplot()*, para o histograma. Para visualizar a soma dos camundongos segundo linhagens registradas, foi utilizado o método *groupby()*, que tem a função de agrupar os dados de acordo com um parâmetro de interesse; neste caso, foi utilizado a coluna de linhagens. Para o cálculo da média dos animais, foi utilizada a função *mean()* nos dados agrupados. Buscando verificar a presença de valores discrepantes nos dados de nascimento, foi gerado o gráfico do tipo *boxplot* com o método *boxplot()* da biblioteca *seaborn* (Figura 3.4).

Construção da Tabela 4.1, os dados foram apresentados através da Frequência Absoluta, sendo calculada a soma total desses valores e a Frequência Relativa (apresentado entre parênteses), foi calculado a partir da divisão do valor de nascimento/ano pela soma total de nascimento anualmente. Na *DataFrame* que continha os dados de nascimento, foi filtrada na coluna 'date_hoje' contendo datas de cada ano para os registros. Para calcular o número de nascimentos, foi utilizado o método *resample()* e informado o parâmetro 'A' que corresponde aos anos, retornando a soma dos nascimentos para cada ano registrado. Para as linhagens, cada nome foi guardado em uma variável do tipo lista e filtrado através do método *isin()* (Figura 3.4).

```
#Soma de animais por linhagem/barras

sns_plot_tot=
nume_ani.groupby('ling').sum().sort_values(by='num_ani',ascending=False) #
agrupando os valores e organizando em ordem decrescente.

plt.figure(figsize=(20,5))
sns.barplot(x='ling', #linhagens
            y='num_ani', #número de animais
            palette="rocket", #cor do gráfico
            data=sns_plot_tot).set_xticklabels(rotation=45, #rotação
            horizontalalignment='right',
            labels=sns_plot_tot['ling'])
```



```

#Soma de animais por linhagem
nume_ani['num_ani'].sum() #Somando o número total de animais

len(nume_ani) #Verificando a quantidade de entradas
#Gerando o gráfico histograma

plt.figure(figsize=(15,10)) #Tamanho do gráfico

plot_his = sns.distplot(nume_ani['num_ani'], hist = True,
                        kde = True,
                        color='red',
                        label = 'N° de animais')

#parâmetros para configurações do gráfico

#Valores discrepantes do número de nascimentos de animais - Dados com
outliers

plt.figure(figsize=(15,15)) #tamanho do gráfico
sns.boxplot( y='ling',
            x='num_ani',
            data=nume_ani.sort_values(by='ling', ascending=False)
) #passando os parâmetros do gráfico

#CONTRUÇÃO DA TABELA 6.1
#Soma dos valores de produção animal para cada ano/ todos os dados/freq
abs
dist = nume_ani.resample('A').sum()
Dist

#Soma dos valores de produção animal para cada ano/ cada linhagem/fre abs

ling = ['Swiss-2 E']
test_lings = nume_ani[nume_ani['ling'].isin(ling)]
test_lings.resample('A').sum()

```

Figura 3.4 - Algoritmo para gerar gráficos do tipo barras, histograma e boxplot a partir da soma total do número de camundongos e por linhagem. Visualização da distribuição dos dados de nascimento.

Para uma análise descritiva do conjunto de dados de nascimento, foi criado a Tabela 4.2, apresentando valores a partir do cálculo de desvio padrão, média, valores de 1º quartil (Q1 - 25%), 2º quartil (Q2 - mediana), 3º quartil (Q3 - 75%) e frequência em que o dado aparece. Em *Python*, foi utilizado o método *describe()* o qual retornou à descrição desses valores. Para o cálculo da distribuição de frequência, os dados foram somados e dividido pelo número total de animais. Para identificar valores de cada linhagem, foi criada uma variável do tipo lista com o nome da linhagem para buscar entre os dados apenas informações da linhagem de interesse, em seguida foi aplicado novamente o método *describe()* Figura 3.5.

```
#Construção da tabela 6.2
#Todo conjunto de dados
nume_ani.describe()

#distribuição de frequencia
nume_ani['num_ani'].sum()/nume_ani['num_ani'].sum()

#cada linhagem
ling = ['AG129']
test_lings = nume_ani[nume_ani['ling'].isin(ling)]
test_lings.describe()

#distribuição de frequência
Test_lings['num_ani'].sum()/nume_ani['num_ani'].sum()
```

Figura 3.5 - Algoritmo para identificação dos valores de produção animal.

Para retirar os dados discrepantes, foi filtrado valores que estivessem acima ou abaixo dos limites inferior e superior dos quartis. Em *Python*, os valores discrepantes foram removidos para obter uma população de acordo com o que é observado na rotina e na literatura referente a criação animal. Para tanto, foi realizado o cálculo entre os quartis pelo método *quantile()* a partir da biblioteca *Numpy*. Esse novo conjunto de dados foi armazenado em uma nova *DataFrame* o que foi então usada nos próximos modelos de análise (Figura 3.6). A Tabela 4.3 contém os valores

recalculados após a retirada dos valores discrepantes. A construção da lógica para obter esse conjunto de informações foi a mesma utilizada para criar a Tabela 4.2. Em *Python*, os códigos são os mesmos, porém, agora utilizando a *DataFrame* contendo os valores sem os *outliers*.

Após realizar a retirada dos *outliers*, também foi alterado os valores de produção animal demonstrados na Tabela 4.1, tendo novamente seus dados recalculados, utilizando a mesma lógica presente no tópico 3.5.3, e apresentados na Tabela 4.4.

```
#Criando uma nova Dataframe, filtrando os valores discrepantes para todo
o conjunto de dados
q1 = nume_ani['num_ani'].quantile(0.25) #valores de 25%
q3 = nume_ani['num_ani'].quantile(0.75) #valores de 75%
iqr = q3 - q1 #determina outlier

sem_out = nume_ani[(nume_ani['num_ani'] > (q1 - 1.5 * iqr)) &
(num_ani['num_ani'] < (q3 + 1.5 * iqr))]
sem_out.head()
#Apresenta as 5 primeiras entradas

#CONTRUÇÃO DA TABELA 6.3
#Todo conjunto de dados
nume_ani.describe()

#distribuição de frequência
sem_out['num_ani'].sum()/sem_out['num_ani'].sum()

#cada linhagem
ling = ['AG129']
test_lings = sem_out[nume_ani['ling'].isin(ling)]
test_lings.describe()

#distribuição de frequência
Test_lings['num_ani'].sum()/sem_out['num_ani'].sum()

#CONTRUÇÃO DA TABELA 6.4
#Soma dos valores de produção animal para cada ano/ todos os dados/freq
abs sem out
dist = sem_out.resample('A').sum()
Dist
```

```
#Soma dos valores de produção animal para cada ano/ cada linhagem/fre abs  
sem out  
  
ling = ['Swiss-2 E']  
test_lings = sem_out[sem_out['ling'].isin(ling)]  
test_lings.resample('A').sum()
```

Figura 3.6 - Algoritmo para retirar *outliers* identificados no conjunto de dados entre registros de nascimento. Valores que estavam abaixo ou acima do limite máximo e mínimo encontrado entre os dados foram filtrados a partir do cálculo dos quartis.

3.4.4. Decomposição da Série Temporal para todo o conjunto de dados de nascimentos e para os grupos Isogênico e Heterogênico

Para a apresentação da distribuição de produção animal por intervalos de tempo, os dados contendo informações das datas registradas e do número de animais foram organizadas mensalmente. Foi criada uma tabela que continha uma coluna com as datas e outra coluna com o número mensal acumulado de animais. Para visualização desses dados, foi gerado um gráfico de linha, o qual apresentou o número total de animais distribuído para cada ano, nesse caso, de 2014 a 2019.

Para determinar a Série Temporal em *Python*, foi necessário a criação de *DataFrame* contendo informações de datas de registros como novo índice. Nesse *DataFrame*, foram mantidas as colunas da data em que foi realizado o registro dos nascimentos, as linhagens e número de animais. A partir desse *DataFrame*, foi utilizado o método *resample()* que retorna o agrupamento das datas dos registros, podendo ser analisado por dia, semanas, meses, etc. Neste caso, foi utilizado o parâmetro 'M', agrupando mês a mês. Junto a ele foi passada função *sum()* para calcular a soma do número de animais nascidos mensalmente (Figura 3.7).

Os dados analisados na série, foram resumidos e descritos na Tabela 4.5, apresentando o número de nascimentos por mês de cada ano de todos os dados, para o grupo isogênico (Tabela 4.6) e heterogênico (Tabela 4.7). Para identificar os valores totais, foi utilizado o método *resample()*, que retornou a soma de nascimentos para cada um dos meses registrados. Já para identificar os valores dos grupos isogênico e heterogênico, as linhagens correspondentes de cada grupo foram

guardadas dentro de uma variável do tipo lista e depois filtrados utilizando o método `isin ()` em todo o conjunto de dados, para filtrar apenas os dados das linhagens presentes na lista (Figura 3.7).

```
#Criando um Dataframe pra plotar minhas séries temporais:

serie_temp_ling = pd.DataFrame(sem_out['date_hoje']) #data de registro do
nascimento
serie_temp_ling['ling'] = sem_out['ling'] #linhagens
serie_temp_ling['num_ani'] = sem_out['num_ani'] #número de animais nasc
serie_temp_ling = serie_temp_ling.set_index('date_hoje') #identificando a
nova coluna como index data_hoje
serie_temp_ling.head()

#Série temporal somando o número de animais de todos os meses por ano
from statsmodels.tsa.seasonal import seasonal_decompose

tempo_mes = serie_temp_ling.resample('M').sum() #'M' para mes

#CONSTRUÇÃO DA TABELA 6.5-7 - series
#Soma dos valores de produção animal para cada ano/ todos os dados/freq
abs sem out
dist = sem_out.resample('M').sum()
Dist

#Lista das linhagens isogênicas
ling_iso = ['Balb/c', 'S129KO', 'C57BL/6', '129/SVE', 'C3H', 'DBA/1J', 'AG129']

serie_lings_i = serie_temp_ling[serie_temp_ling['ling'].isin(ling_iso)]
serie_lings_i.resample('M').sum()
#####
#Lista de linhagens heterogênicas
ling_het = ['Swiss-1 A', 'Swiss-1 B', 'Swiss-1 C', 'Swiss-1 D', 'Swiss-1 E',
           'Swiss-2 A', 'Swiss-2 B', 'Swiss-2 C', 'Swiss-2 D', 'Swiss-2 E']
serie_lings_h = serie_temp_ling[serie_temp_ling['ling'].isin(ling_het)]
serie_lings_h.resample('M').sum()
```

Figura 3.7 - Determinação de *DataFrame* para construção da série temporal a partir dos dados de nascimento de animais. Importação de bibliotecas para decomposição da série temporal.

Para decompor as séries, foi necessário importar da biblioteca *Statsmodels* a classe *tsa* (*Time Series Analysis*). Foi utilizado método da classe *tsa seasonal_decompose()*, com o parâmetro de 12 meses, para decompor os dados a cada ano, de 2014 a 2019. Com esta função, os resultados são obtidos estimando primeiro a tendência, aplicando o filtro de convolução aos dados a cada 12 meses. A tendência é então removida da série e a média dessa série descendente para cada período é o componente sazonal retornado. A plotagem desses dados foi separada em quatro gráficos diferentes. O primeiro gráfico corresponde ao número de animais produzidos mês a mês e foi representado a cada ano dentro do período analisado. Os demais são os componentes da série, tendência, sazonalidade e aleatoriedade (Figura 3.8).

Para gerar a decomposição por grupos de linhagens isogênicas ou heterogênicas, foi necessário filtrar as linhagens entre os dois grupos, utilizando a mesma lógica apresentada na Figura 3.7

```
#agrupando por mês e somando número de animais
decom = seasonal_decompose(tempo_mes, freq=12)
#decompondo a série temporal

+++++
# Plotagem de gráfico

plt.figure(figsize=(20, 10)) #tamanho do gráfico

plt.subplot(4, 1, 1)
plt.plot(tempo_mes, label = 'N° nascidos', color='red')
plt.rcParams.update({'font.size': 15}) #tamanho da fonte
#gráfico com o número de animais nascidos

plt.title('Decomposição da Série Temporal: Nascimento de Animais Somando
os Meses no período de 5 anos.')
plt.legend(loc = 'best') #melhor local para legenda

plt.subplot(4, 1, 2)
plt.plot(decom.trend, label = 'Tendência', color='green')
plt.legend(loc = 'best') #melhor local para legenda
```

```

#gráfico plotando a tendência

plt.subplot(4, 1, 3)
plt.plot(decom.seasonal, label = 'Sazonalidade', color='blue')
plt.legend(loc = 'best') #melhor local para legenda
#gráfico plotando a sazonalidade

plt.subplot(4, 1, 4)
plt.plot(decom.resid, label = 'Ruído', color='grey')
plt.legend(loc = 'best') #melhor local para legenda
#gráfico plotando a aleatoriedade

plt.xlabel('Data Nascimento')
plt.tight_layout()

```

Figura 3.8 – Algoritmo de decomposição da série temporal a partir dos dados mensais de nascimento de camundongos para todo conjunto de dados. Grupos Isogênico e Heterogênico. Decomposição da tendência, sazonalidade e ruído.

3.4.5. Mediana da produção de camundongos para os grupos Isogênicos e Heterogênicos

Dada a constituição genética dos grupos isogênico e heterogênico, é esperado que o número de nascimentos seja diferente entre eles, resultando em maior prole para grupo heterogênico (constituição genética variada) comparada ao grupo isogênico (menor variação genética). Para identificar esses valores, conhecendo assim quantos animais foram produzidos para cada grupo, calculamos a mediana de nascimento para os grupos isogênico e heterogênico. Para isso, foi necessário trabalhar novamente com as listas contendo as linhagens para cada grupo, como visto nos tópicos anteriores. Filtrando apenas as linhagens (da coluna 'ling') do grupo isogênico ou heterogênico, foi calculado a mediana dos valores da coluna 'num_aní' apresentados em um gráfico de barras. Para gerar a linha de média e mediana foi calculado a média e mediana a partir dos valores medianos de nascimento das linhagens.

Para determinar a mediana de nascimentos nos grupos de diferentes antecedentes genéticos, foram utilizadas as listas criadas anteriormente na decomposição em série. Gráficos separados foram gerados para analisar a mediana

dos valores usando *median()* para cada linhagem dentro de cada grupo, novamente usando os métodos *plot()* e *groupby()* para agrupar por linhagens. Para traçar os intervalos médios e medianos, esses valores foram calculados novamente com base nos nascimentos medianos. Para gerar o gráfico, foi utilizado a biblioteca *Seaborn* e *Matplotlib* (Figura 3.9).

```
#Gráfico-calcula da mediana de animais nascidos
#isogênico
g_iso =
sem_out[sem_out['ling'].isin(ling_iso)].groupby('ling')
linhagens_iso =
g_iso.median()

#heterogênico
g_het =
sem_out[sem_out['ling'].isin(ling_het)].groupby('ling')
linhagens_het =
g_het.median()

plt.figure(figsize=(20,5))
sns.barplot(x='ling', #linhagens y='num_ani', #número de
animais
palette="rocket", #cor data=sns_plot_iso).set_xticklabels( rotation=45,
horizontalalignment='right', labels=sns_plot_iso['ling'])
```

Figura 3.9 - Algoritmo com valores medianos de nascimento para os grupos Isogênico e Heterogênico. Apresentado em gráfico de barras.

3.4.6. Mediana do intervalo entre partos (IEP) e do nascimento ao desmame

Na produção animal, alguns parâmetros de criação de camundongos já são bem definidos. Ao atingir uma idade entre 6-8 semanas (idade reprodutiva), os camundongos (macho e fêmea) são colocados em acasalamento, onde permanecem juntos até o nascimento da prole. Para analisar o tempo de nascimento dos animais, foi considerado o intervalo de tempo do momento no qual os camundongos são colocados em acasalamento até o nascimento da ninhada. Foi calculado a diferença dos registros de datas da coluna 'data_acas', com datas da montagem do acasalamento, com a coluna 'data_nasc_acas', contendo as datas de nascimento. Os resultados foram guardados em uma variável e em seguida foi criada uma coluna contendo a diferença entre esses valores.

Assim que os animais atinjam a idade de desmame e estejam comendo ração sólida, estes são separados das mães e remanejados para novas caixas de criação. Para identificar o período do desmame, a mesma lógica foi aplicada, calculando-se a diferença entre a data de nascimento e o momento do desmame (sexagem). Neste caso, foi calculada a diferença entre os registros de datas da coluna 'data_nasc_acas' com a coluna 'date_sex'.

Dentre os registros de intervalo entre partos, foram somadas 1.799 entradas (registro de quantas vezes os casais foram colocados em acasalamento) e para calcular a mediana desse período, foi necessário criar uma *DataFrame* para incluir uma nova coluna contendo o número de dias entre o registro acasalamento até o nascimento dos animais. Foi utilizado o método *lambda* que permite criar uma função que utilizada apenas uma vez, nesse caso, a variável 'day_nasc_acas', que recebe a subtração entre a coluna 'date_nasc_acas' e 'date_acas', que representa o número de dias entre o registro de acasalamento e nascimento. Entre os registros do desmame, foram somadas 4.938 entradas (registro de quantas vezes os filhotes foram sexados e separos da mãe) analisadas, e para encontrar esse intervalo, foi utilizado as colunas 'date_nasc_acas' e 'date_sex' (Figura 3.10).

```
#Criada um DataFrame com o período entre acasalamento e nascimento - com
outliers (as datas day_nasc_acas)

nume_ani_nasc_acas = pd.DataFrame()
#Criando um novo DataFrame
nume_ani_nasc_acas = sem_out[['date_acas', 'date_nasc_acas', 'num_nasc',
'ling']]
#incluindo as colunas que eu preciso
nume_ani_nasc_acas['day_nasc_acas'] = list(map(lambda day_nasc_acas:
sem_out['date_nasc_acas'] - sem_out['date_acas'], sem_out))[0]
nume_ani_nasc_acas.head()
#subtraindo os valores da date_nasc_acas por date_acas e resultar no
valor dos dias de acasalamento e nascimento

#*****
#Gráfico de acasalamento ao nascimento
```

```

sns_plot_aca =
sem_out_acas.groupby('ling').median().sort_values(by='day_nasc_acas_day',
ascending=False)

plt.figure(figsize=(15,10))
sns.barplot(x='ling', #Linhagens
            y='day_nasc_acas_day', #coluna com a diferença de dias
            palette="rocket", #cor
            data=sns_plot_aca).set_xticklabels(
            rotation=45,
horizontalalignment='right',
            labels=sns_plot_aca['ling'])

```

Figura 3. 10 - Algoritmo para calcular o intervalo entre partos e do nascimento ao desmame para todas as linhagens. Apresentado em gráfico de barras.

3.4.7. Linhagens solicitadas com mais frequência por usuários internos e externos

O LACEA produz diferentes linhagens de camundongos para atender solicitações entre os pesquisadores internos do Instituto Carlos Chagas e solicitações de Instituições externas. No BioterC, existe uma sessão de solicitação utilizada pelos clientes, onde o usuário informa ao sistema quantos animais vai precisar, qual linhagem, define se é macho ou fêmea, qual idade desses animais e o número do Comitê de Ética (CEUA). A partir dessas informações, os bioteristas apresentam uma previsão da entrega desses animais, considerando as etapas de criação. Dentro do BioterC, esses dados de solicitação foram armazenados na tabela 'pedido' (Tabela 3.2 - Materiais e Métodos).

Para buscar as linhagens mais solicitadas nos dados, foi necessário importar a tabela SQL "pedido", a qual é formado pelas variáveis descritas na Tabela 3.2., a qual apresenta um total de 403 entradas. A Tabela 4.8 descreve o conteúdo da tabela "pedido" onde são apresentados o número de solicitações registradas, e o número de animais solicitados anualmente. Para encontrar esses valores, foi utilizado o método *resample()*, retornando a soma de animais solicitados anualmente (Frequência Absoluta). O cálculo da Frequência Relativa, foi através da soma de animais solicitados a cada ano e dividido pelo número total de animais solicitados no período.

Tabela 3.2 - Informações das colunas utilizadas para análise dos dados a partir das solicitações, armazenadas na tabela pedido.

Colunas da Tabela	Descrição do que contém em cada coluna BioterC
'solic'	Nome do solicitante.
'lab'	Qual laboratório.
'ling'	Linhagens.
'num_ani'	Número de animais solicitados.
'cep'	Registro do Comitê de Ética.
'data_inicio'	Data de início da experimentação.
'data_solic'	Data de solicitação do animal.
'data_forn'	Data de fornecimento.
'data_fim'	Data da finalização da experimentação.

403 entradas x 09 colunas

Fonte: BioterC, 2019

Para identificar as linhagens mais solicitadas entre usuários internos e externos, foi filtrado primeiramente os registros de usuários internos da coluna que contém informações dos laboratórios ('lab'). Uma vez determinado os usuários internos, foi somado os valores de animais solicitados para cada linhagem. Ao agrupar esses valores por linhagem, os dados nos retornaram a soma de cada registro em ordem do maior para o menor, podendo assim, identificar as linhagens mais solicitadas. A mesma lógica foi aplicada aos usuários externos.

Para representar todas as linhagens solicitadas entre usuários internos e externos, foi criada a Tabela 4.9, apresentando registros de solicitação de cada linhagem solicitada entre os laboratórios internos e instituições externas. Para o cálculo de Frequência Absoluta foi utilizado o método *groupby()*, onde os valores foram agrupados por linhagem e o número de animais solicitados foi somado. Para o cálculo da Frequência Relativa foi feita a divisão do valor somado de cada linhagem pelo número total entre as solicitações (Figura 3.11).

Para facilitar nossa análise utilizando *Python*, esses dados foram filtrados primeiro apenas para informações de usuários internos a partir da coluna 'lab', e um *DataFrame* foi criado a partir desses dados. Os dados foram agrupados por linhagens e os valores somados pela função *sum()*, para facilitar a identificação das linhagens

mais solicitadas. Para pesquisar as linhagens mais solicitadas por instituições externas, foi utilizada a mesma lógica (Figura 3.11).

```
#Criando uma variável para receber o filtro só de usuários internos

pedidos_int = pedido[pedido['lab'] != 'Usuário Externo'] #usando filtro a partir de "!=" que significa diferente de. Me apresentando apenas os pedidos internos
pedidos_int.head()

#####
#Transformando todas as datas que estão em segundos para formato de data

datas = ['data_inicio', 'data_solic', 'data_forn', 'data_fim'] #datas registradas

for d in datas:
    pedidos_int[d] = pd.to_datetime(pedidos_int[d], unit = 's') #indica o formato das datas
#####
pedidos_int.head()
inte = pedidos_int.groupby('ling') #usando groupby para agrupar
inte.sum() #somando o total de animais por linhagem

#####
ling_mais_soli_ano = ['Balb/c', 'Swiss', 'S129KO'] #lista com + solic

g_ling_mais_ped_in =
pedidos_int[pedidos_int['ling'].isin(ling_mais_soli_ano)].groupby('ling')
lin_mais_ped_in = g_ling_mais_ped_in.sum()
lin_mais_ped_in
```

```
#CONTRUÇÃO DA TABELA 6.8
#Soma dos valores de solicitação de camun para cada ano/ todos os dados/freq abs sem out
dist = pedido_int.resample('A').sum()
Dist
```

```

#Soma dos valores de produção animal para cada ano/ cada linhagem/fre abs
sem out
#CONTRUÇÃO DA TABELA 6.9
pedidos_int.head()
lings= pedidos_int.groupby('ling') #usando groupby para agrupar
lings.sum() #somando o total de animais por linhagem

```

Figura 3. 11 - Algoritmo para buscar nos dados, quais as linhagens dos grupos isogênico e heterogênico mais solicitadas entre instituições internas e externas. Gráfico apresentando as três linhagens mais solicitadas entre os usuários e instituições externas que solicitaram com mais frequência.

3.4.8. Instituições externas que mais solicitaram camundongos

Entre os dados de solicitações, foram identificadas diferentes instituições externas. Como desdobramento dos resultados anteriores, buscamos quais delas mais solicitaram animais ao LACEA. Uma vez determinada as linhagens, para trabalhar com os dados de solicitantes externos, dessa vez foi filtrada a coluna 'extern' com o registro dos nomes das instituições e somado o número de entradas (registros) para cada um dos dados, e posteriormente somado os valores de cada entrada.

Em *Python*, para buscar entre as instituições as que mais solicitaram camundongos no período avaliado, foi necessário filtrar a coluna que contém as informações dos solicitantes 'extern' e, para identificar os solicitantes, foi utilizado a função *values.counts()* que retornou as instituições e o número de entradas para cada uma delas, apresentando valores do maior para o menor. Esses dados foram agrupados pela função *groupby()* e depois somado os valores através de *sum()* (Figura 3.12).

Para representar os registros das instituições externas, foi criada a Tabela 4.10, apresentando registros de solicitação entre os usuários externos. O cálculo de Frequência Absoluta foi feito através do método *groupby()*, onde os valores foram agrupados por instituições e o cálculo realizado foi através da mesma lógica para construir a Tabela 4.9 (Figura 3.11).

```

#Filtrando a coluna das instituições externas 25 primeiras
pedidos_ext['exter'].value_counts()[:25] #retorna insti e n° de entrada

sns_plot_ped_ins =
pedidos_ext.groupby('exter').sum().sort_values(by='num_ani',ascending=False).reset_index()
sns_plot_ped_ins

plt.figure(figsize=(20,5))

sns.barplot(x='exter',
            y='num_ani', palette="rocket",
            data=sns_plot_ped_ins).set_xticklabels(
                rotation=45,
                horizontalalignment='right',
                labels=sns_plot_ped_ins['exter']
            )

plt.legend([str(plot_media) + ' mean mouse/strain', str(plot_mediana) + '
median mouse/strain'])
plt.xlabel('Instituições Externas') #definindo nome do eixo X
plt.ylabel('Número de Animais') #definindo nome do eixo Y

```

Figura 3. 12- Algoritmo para gerar gráficos apresentando solicitação de camundongos entre usuários externos. Os dados foram agrupados e plotados em um gráfico de barras.

3.4.9. Motivos de descarte dos camundongos em todo o conjunto de dados, para os grupos Isogênicos e Heterogênicos e para a linhagem C57BL/6

A última etapa presente na criação animal é o descarte. É prática comum entre os biotérios realizar o descarte de animais em diferentes situações, sendo este um indicador que pode trazer informações sobre custos a instituição devido a manutenção de animais em estoque. Para qualificar e quantificar o motivo de descarte dos camundongos, foi necessário verificar na tabela “est_est_log” a variável “obs_desc” a qual contém o motivo de descarte dos animais. Uma vez identificado o motivo de descarte, a primeira estratégia foi buscar nessa coluna os motivos que retornassem maior número de registros. Por se tratar de um campo aberto, foram identificados motivos correspondentes (sinônimos), porém, escritos de outra forma,

por exemplo, "Descarte Zootécnico" era sinônimo de "descarte Zootécnico" ou "descarte zootécnico - briga". Para permitir as análises, cada motivo que se repetisse entre os dados foi agrupado em uma única representação, até que fosse obtido pelo menos nove motivos diferentes: Descarte Zootécnico, Fim Experimento, Encontrado Morto, Velhos, Canibalismo, Fornecido Zoológico, Animais Doentes, Monitoramento Sanitário, Ivermectina. Dados que não correspondiam a esses principais motivos foram agrupados em uma décima categoria denominada de 'Outros'. Após esse processamento os dados foram filtrados para cada linhagem, separadas pelo grupo isogênico e heterogênico e identificados os motivos de descarte mais frequentes em cada grupo. Os motivos de descarte de machos e fêmeas da linhagem C57BL/6 (linhagem com a maior produção do LACEA) foram ainda, analisados separadamente.

Para obter este conjunto de dados de descarte da tabela 'est_est_log' foi filtrado o valor 'desc' da variável 'útil' o qual representa os dados de descarte, e a coluna 'obs_desc', onde está presente os registros do motivo de descarte, procedimento que retornou 2.286 eventos únicos. Por apresentar um campo aberto para digitação, foi necessário processar e minerar os dados. Para a estruturação e processamento dos dados, foi identificado todos os motivos de descarte únicos que já foram registrados através da função *value_counts()*, a qual retornou os dados e sua frequência absoluta. Entre esses dados, foram identificados os termos que retornavam pelo menos 10 ou mais entradas com o mesmo nome. Os demais registros foram concatenados e somados aos termos que apareciam com mais frequência através da função *replace()*. Após essa triagem, os dados ainda indicavam diferentes motivos para o descarte; 90% dos dados foram separados e outros 10% foram concatenados e renomeados como "Outros". Para a realização do cálculo, foi utilizada novamente a função *groupby()*, mas, neste caso, os dados foram agrupados pela coluna dos motivos do descarte e os valores foram somados pela função *sum()*. Para os grupos isogênicos e heterogênicos foi filtrada as linhagens armazenadas na lista citada no tópico 5.5.4. Para a linhagem C57BL/6, foi necessário criar um *DataFrame*, filtrando as informações apenas desta linhagem e agrupando-as novamente entre os dados de machos e fêmeas que foram analisados separadamente. Para visualizar melhor os dados, nós os plotamos em um gráfico de barras (Figura 3.13).

Para ter conhecimento dos dados entre os descartes, foi criada a Tabela 4.11 contendo os valores de descartes registrados, nº de animais descartados,

apresentando a Frequência Absoluta e a Frequência Relativa para cada ano registrado. Para encontrar esses valores, foi utilizada a mesma lógica para construir a Tabela 4.9 apresentado no tópico 3.5.7.

Para construção das Tabelas 4.12-14, foi encontrado o número de animais descartados para cada motivo de descarte, apresentando Frequência Absoluta que foi somado utilizando o método `groupby()`, retornando os valores somados de descarte para cada motivo e, para calcular a Frequência Relativa, foi dividido os valores de cada motivo pelo número total de animais descartados. Para as Tabelas 4.15-16, primeiro foi necessário filtrar apenas registros da linhagem C57BL/6, aplicado a mesma lógica utilizada para criar as Tabelas 4.12-14.

Para construir a tabela 4.17 com os motivos de descarte animal que não correspondiam a descartes comuns entre os biotérios, esses dados foram filtrados utilizando o método `groupby()` e para o cálculo da Frequência Absoluta e Relativa foi aplicado a mesma lógica utilizado para criar as Tabelas 4.12-14.

```
#Lista que vai guardar meus 90% dos dados a serem analisados

lista_10por_super = list(descarte['obs_desc'].value_counts()[:18].index)
#18 primeiras entradas
lista_10por_infer = list(descarte['obs_desc'].value_counts()[18:].index)
#18 últimas entradas

#Gráfico contendo motivos de descarte
sns_desc_tot =
descarte_filtrada.groupby('obs_desc').sum().sort_values(by='num_ani', ascending=False).reset_index()
sns_desc_tot
#+++++
#Para grupo Isogênico

g_iso =
descarte_filtrada[descarte_filtrada['ling'].isin(ling_iso)].groupby('ling')
g_iso.sum()

#+++++
#Para grupo heterogênico
```



```

g_het =
descarte_filtrada[descarte_filtrada['ling'].isin(ling_het)].groupby('ling')
g_het.sum()

#####
#Para linhagem C57BL/6

linhagem_macho = linhagem_black[linhagem_black['m_f'] == 'm'] #Filtrando
machos

linhagem_macho = linhagem_black[linhagem_black['m_f'] == 'f'] #Filtrando
fêmeas

#Construção das tabelas 6.12-18 - todas foram utilizando método groupby

descarte_filtrada.groupby('obs_desc').sum() #freq. Abs.

#####
#Isogênicos e Heterogênicos

descarte_filtrada[descarte_filtrada['ling'].isin(ling_iso)].groupby('ling')
g_iso.sum() #freq. Abs.

descarte_filtrada[descarte_filtrada['ling'].isin(ling_het)].groupby('ling')
g_het.sum() #freq. Abs.

```

Figura 3. 13 - Algoritmo para processamento e mineração dos motivos de descarte animal a partir dos registros de nascimento. Dados apresentados em gráficos e tabelas.

4. Resultados

4.1. Conjunto de dados do BioterC apresenta valores discrepantes para nascimento de animais

Em cinco anos, foram produzidos pelo LACEA e registrados no BioterC 15.106 camundongos em 1.808 registros (entradas) (Tabela 4.1). A distribuição dos dados de nascimentos está representada em histograma Figura 4.1 A, que retornou registros de nascimento entre valores de 01 até 60 animais por evento de nascimento. Valores acima de 30 animais por nascimento foram considerados errados, levando em consideração a experiência dos bioteristas do LACEA e a bibliografia. Na soma total dos animais nascidos nesse período, para cada linhagem (Figura 4.1 B e Tabela 4.1), os resultados demonstram que, as linhagens mais produzidas foram Balb/c, (4.289 animais, 28%), C57BL/6 (4.001 animais, 26%) e Swiss (3.395 nascimentos, 22%). É pertinente salientar que, como grupo heterogêneo, a linhagem Swiss é dividida em dez subgrupos (Swiss-1A – 1E e Swis-2A - 2E), apresentando valores abaixo ou próximos da mediana (Figura 4.1 B, linha verde), porém distantes da média (Figura 4.1 B, linha vermelha). A soma dos valores de nascimento para as demais linhagens está apresentada na Tabela 4.1. A discrepância observada entre o valor médio (888,59) e mediano (381,0) indica que os dados não estão normalmente distribuídos, o que reforça a hipótese da presença de *outliers* na população.

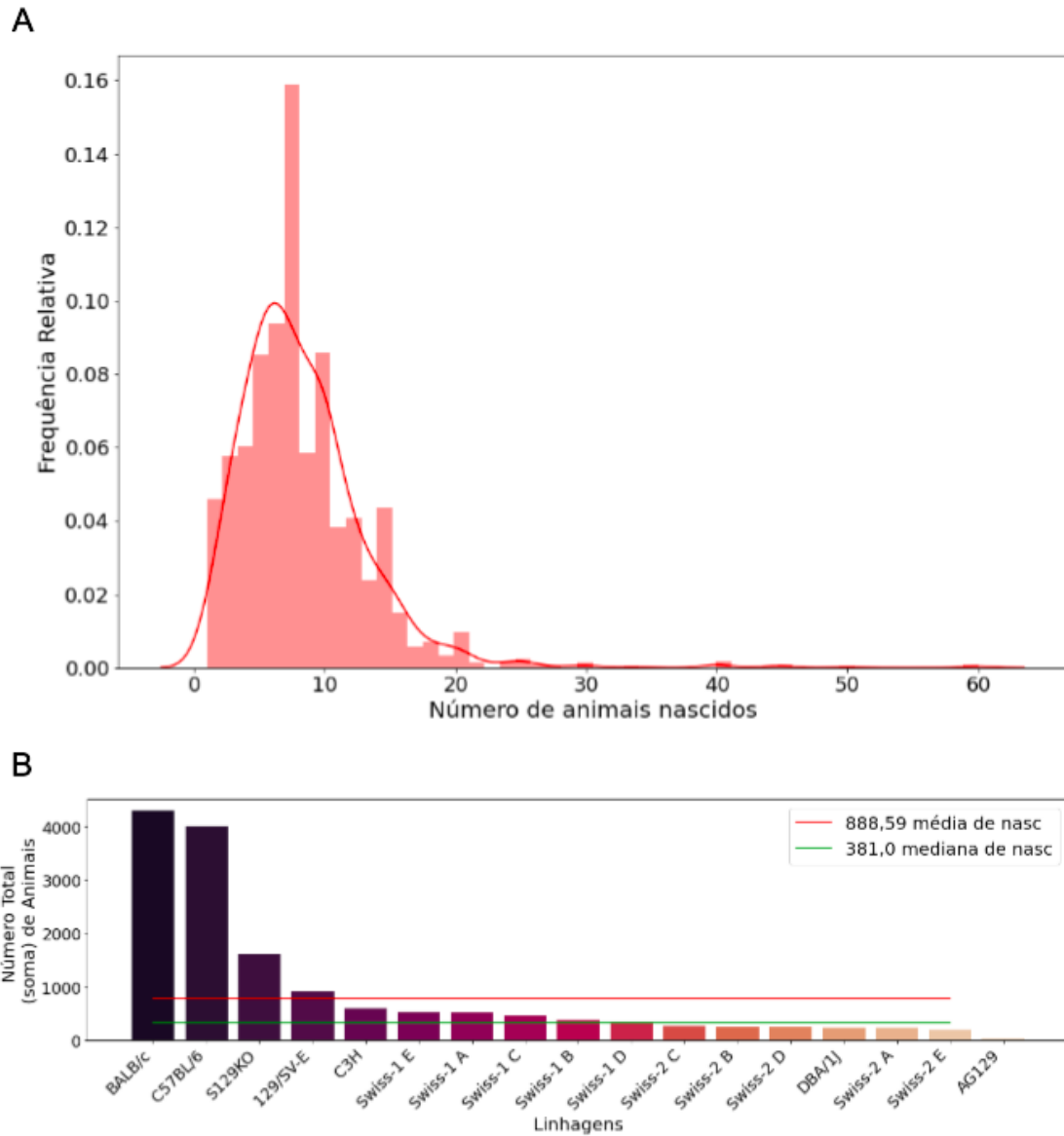


Figura 4.1 - Distribuição do número de animais a partir dos registros de nascimento (A) Representação por histograma da distribuição de nascimentos de camundongos por evento de nascimento. **(B)** Soma total de camundongos registrados no BioterC separados por cada linhagem produzida.

Tabela 4.1 - Distribuição do número de animais considerando como unidade observacional o registro de produção anual e o número de nascimentos (total e para cada linhagem).

ANO	2014	2015	2016	2017	2018	2019	TOTAL
N° de entradas	52 (0,02)	196 (0,10)	357 (0,19)	328 (0,18)	559 (0,30)	316 (0,17)	1.808
Dados de Produção (N° de nascimentos)							
N° de nascimentos (Total)	599 (0,03)	1.629 (0,10)	2.881 (0,19)	2.607 (0,17)	4.562 (0,39)	2.828 (0,18)	15.106
Balb/c	107 (0,02)	774 (0,18)	859 (0,20)	541 (0,12)	1.409 (0,32)	599 (0,13)	4.289 (0,28)
C57BL/6	6 (0,001)	178 (0,04)	397 (0,09)	691 (0,17)	1.645 (0,41)	1.084 (0,27)	4.001 (0,26)
S129KO	14 (0,008)	166 (0,10)	242 (0,15)	318 (0,19)	336 (0,20)	563 (0,34)	1.612 (0,11)
129/SV-E	100 (0,11)	159 (0,17)	297 (0,32)	121 (0,13)	101 (0,11)	127 (0,14)	905 (0,06)
C3H	-	-	203 (0,33)	166 (0,27)	194 (0,32)	20 (0,03)	603 (0,04)
DBA/1J	-	-	-	-	-	230 (1,0)	230 (0,015)
AG129	-	-	-	-	43 (1,0)	-	43 (0,003)
Swiss-1 A	28 (0,05)	83 (0,16)	236 (0,45)	91 (0,17)	67 (0,13)	10 (0,01)	515 (0,03)
Swiss-1 B	24 (0,06)	41 (0,10)	116 (0,02)	96 (0,25)	91 (0,23)	13 (0,03)	381 (0,02)
Swiss-1 C	144 (0,31)	65 (0,14)	107 (0,23)	88 (0,19)	47 (0,10)	11 (0,02)	462 (0,03)
Swiss-1 D	10 (0,03)	83 (0,24)	88 (0,26)	77 (0,23)	54 (0,16)	21 (0,06)	333 (0,02)

Swiss-1 E	166 (0,31)	80 (0,15)	29 (0,05)	104 (0,19)	132 (0,24)	20 (0,03)	531 (0,03)
Swiss-2 A	-	-	81 (0,35)	61 (0,26)	66 (0,29)	18 (0,07)	226 (0,14)
Swiss-2 B	-	-	55 (0,21)	62 (0,24)	85 (0,33)	50 (0,19)	252 (0,017)
Swiss-2 C	-	-	72 (0,27)	75 (0,28)	99 (0,37)	20 (0,07)	266 (0,017)
Swiss-2 D	-	-	52 (0,20)	58 (0,23)	118 (0,47)	21 (0,08)	249 (0,016)
Swiss-2 E	-	-	47 (0,22)	58 (0,27)	75 (0,36)	28 (0,13)	208 (0,014)
Swiss (Soma Subgrupos)	344 (0,57)	352 (0,22)	883 (0,31)	770 (0,29)	834 (0,18)	212 (0,07)	3.395 (0,22)

*Valores exibidos na tabela estão apresentados em frequência absoluta e em entre parênteses frequência relativa.

Fonte: a autora

Buscando identificar a presença de valores discrepantes, os registros de nascimento foram analisados através do gráfico de *boxplot* (que representam as variações de uma observação de dados numéricos por meio de quartis de distribuição), descritas detalhadamente na Tabela 4.2. Ao questionar os bioteristas a respeito do número máximo de nascimento que os camundongos podem apresentar a cada gestação (máximo de 15-18) e com base nas análises, foi possível comprovar a presença de *outliers* (Figura 4.2). A presença de valores, como, por exemplo, de 60 animais por nascimento foi registrada duas vezes, podendo indicar erros de registro de nascimento.

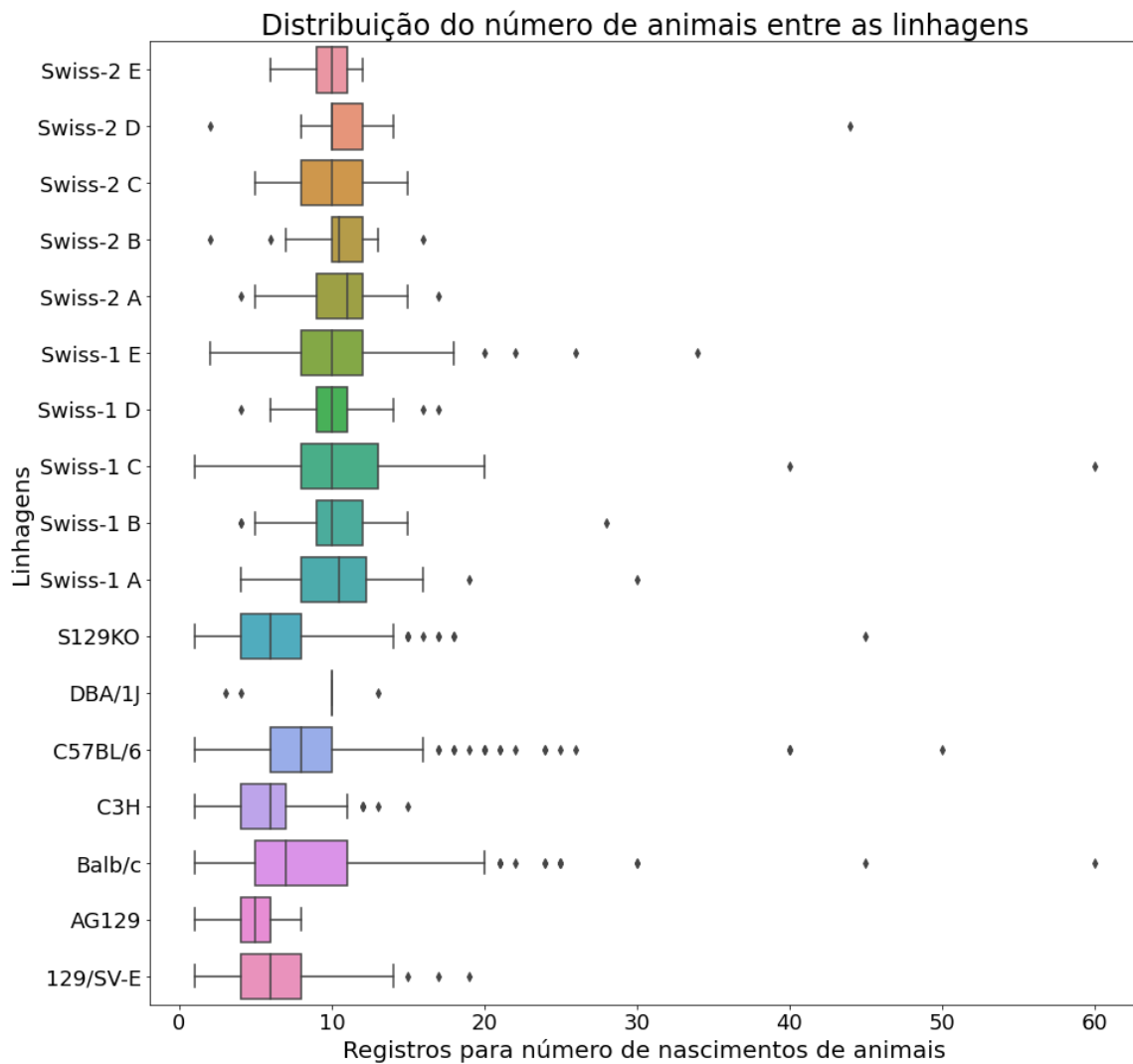


Figura 4.2 - Identificação de valores discrepantes (*outliers*) no conjunto de dados de registros de nascimentos. Gráfico representando a variação de dados observados para o número de animais produzidos. Cada ponto representa uma entrada única no banco de dados de registro de nascimentos.

Tabela 4.2 - Identificação de *outliers* nos registros de nascimento total e para cada linhagem.

	Desvio padrão	Média	Mín	1°Quartil (25%)	2°Quartil (Mediana)	3°Quartil (75%)	Máx	Distr. de freq.
Número de animais produzidos (total)	5,241	8,355	1	5	7	10	60	1.0
Balb/c	5,852	8,612	1	5	7	11	60	0,283
C57BL/6	5,178	8,697	1	6	8	10	50	0,264
S129KO	4,234	6,448	1	4	6	8	45	0,106
129/SV-E	3,042	6,073	1	4	6	8	19	0,059
C3H	2,821	5,911	1	4	6	7	15	0,039
DBA/1J	1,976	9,583	3	10	10	10	13	0,015
AG129	2,166	4,777	1	4	5	6	8	0,002
Swiss-1 A	4,480	10,729	4	8	10,5	12,25	30	0,034
Swiss-1 B	3,974	10,583	4	9	10	12	28	0,025
Swiss-1 C	10,164	12,157	1	8	10	13	60	0,025
Swiss-1 D	2,897	10,090	4	9	10	11	17	0,022
Swiss-1 E	5,755	11,297	2	8	10	12	34	0,035
Swiss-2 A	3,176	10,761	4	9	11	12	17	0,014
Swiss-2 B	2,766	10,500	2	10	10.5	12	16	0,016
Swiss-2 C	2,444	9,851	5	8	10	12	15	0,017
Swiss-2 D	7,786	11,857	2	10	10	12	44	0,016
Swiss-2 E	1,700	9,904	6	9	10	11	12	0,013

Fonte: a autora

Uma vez retirados os valores que se apresentaram além do limite mínimo e máximo dos quartis, foi possível observar uma população mais próxima da distribuição normal, o que melhor representa as observações de nascimento na rotina do LACEA e o que é descrito na literatura. Na representação gráfica em histograma dos dados, sem a presença dos *outliers*, a população apresenta um máximo de 17 camundongos por evento de nascimento (Figura 4.3 A), sendo esses valores sem *outliers* descrito detalhadamente na Tabela 4.3.

Tabela 4.3- Após a identificação dos *outliers*, os valores discrepantes foram retirados e apresentados novamente a distribuição dos valores de nascimento.

	Desvio padrão	Média	Mín	1°Quartil (25%)	2°Quartil (Mediana)	3°Quartil (75%)	Máx	Distr de freq.
Número de animais produzidos (total)	3,614	7,674	1	5	7	10	17	1,0
Balb/c	3,853	7,488	1	5	7	10	17	0,259
C57BL/6	3,457	8,020	1	6	8	10	17	0,266
S129KO	3,311	6,198	1	4	5	8	17	0,114
129/SV-E	2,859	5,986	1	4	6	7,25	17	0,066
C3H	2,821	5,911	1	4	6	7	15	0,045
DBA/1J	1,976	9,583	3	10	10	10	13	0,172
AG129	2,166	4,166	1	4	5	6	8	0,03
Swiss-1 A	3,290	10,130	4	8	10	12	16	0,034
Swiss-1 B	2,660	10,085	4	9	10	12	15	0,026
Swiss-1 C	3,526	9,5	1	8	9,5	11	16	0,024
Swiss-1 D	2,897	10,090	4	9	10	11	17	0,024
Swiss-1 E	3,130	9,585	2	8	10	11	16	0,029
Swiss-2 A	3,176	10,761	4	9	11	12	17	0,016
Swiss-2 B	2,766	10,5	2	10	10,5	12	16	0,018
Swiss-2 C	2,444	9,851	5	8	10	12	15	0,019
Swiss-2 D	2,593	10,25	2	9,75	10	12	14	0,015
Swiss-2 E	1,700	9,904	6	9	10	11	12	0,015

Fonte: a autora

Na ausência dos valores discrepantes, foi novamente analisada a soma de animais por linhagem, buscando-se verificar se houve alguma alteração em relação ao dado anterior. O conjunto de dados agora apresenta total de 13.323 nascimentos em 1.736 registros (entradas). Também, foi possível observar no gráfico da Figura 4.1 B e na Tabela 4.4 que a linhagem Balb/c (agora com 3.452 animais, 26%) era a mais produzida, porém, ao retirar os *outliers* a linhagem C57BL/6 (3.553 animais, 27%) foi

a mais produzida e o grupo Swiss ao ser somado, retorna 3.025 animais, 23%, permanecendo em terceiro lugar como o mais produzido. Os valores de produção para as demais linhagens estão descritos na Figura 4.3 B e na Tabela 4.4.

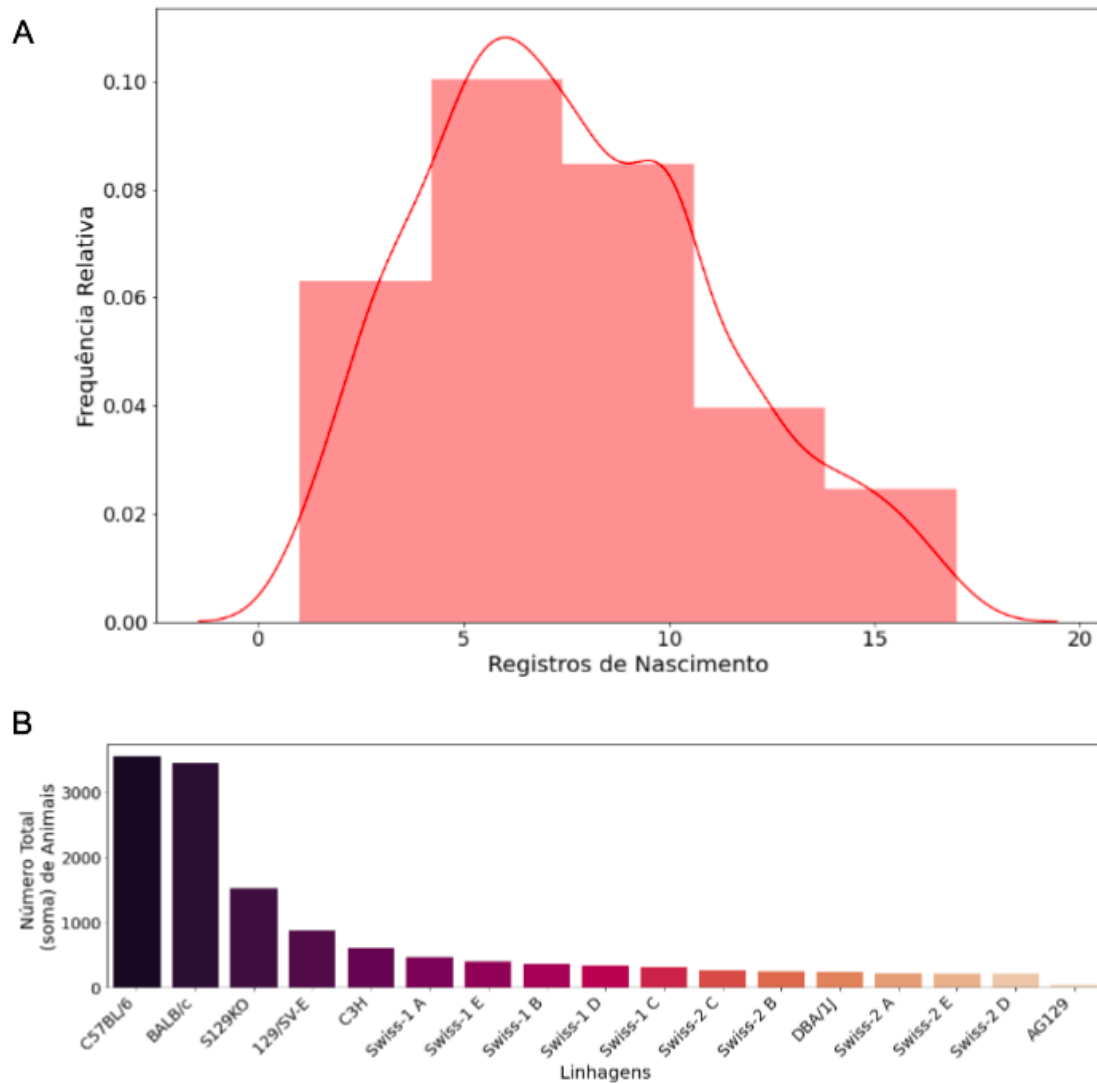


Figura 4.3 - Distribuição do número de animais a partir dos registros de nascimento após a retirada de outliers. (A) Histograma após a retirada dos outliers apresenta uma curva simétrica na distribuição dos dados. (B) Após a retirada dos valores discrepantes, as linhagens C57BL/6 e BALB/c foram mais frequentemente produzidas, com uma diferença de 101 camundongos nascidos.

Tabela 4.4- Distribuição do número de animais considerando como unidade observacional o registro de produção anual e o número de nascimentos (total e para cada linhagem sem a presença de *outliers*).

ANO	2014	2015	2016	2017	2018	2019	TOTAL
N° de entradas	42 (0,02)	188 (0,10)	336 (0,19)	323 (0,18)	547 (0,31)	300 (0,17)	1.736
Dados de Produção							
(N° de nascimentos)							
N° de nascimentos (Total)	322 (0,02)	1.378 (0,10)	2.507 (0,18)	2.482 (0,18)	4.239 (0,31)	2.395 (0,17)	13.323
Balb/c	89 (0,02)	615 (0,17)	553 (0,16)	505 (0,14)	1.210 (0,35)	480 (0,13)	3.452 (0,26)
C57BL/6	6 (0,001)	154 (0,04)	357 (0,10)	602 (0,16)	1,583 (0,44)	851 (0,23)	3.553 (0,27)
S129KO	14 (0,009)	166 (0,10)	242 (0,15)	318 (0,20)	336 (0,21)	455 (0,29)	1.531 (0,11)
129/SV-E	81 (0,09)	159 (0,17)	297 (0,33)	121 (0,13)	101 (0,11)	127 (0,14)	886 (0,07)
C3H	-	-	203 (0,33)	166 (0,27)	194 (0,32)	40 (0,06)	603 (0,05)
DBA/1J	-	-	-	-	-	230 (1,0)	230 (0,02)
AG129	-	-	-	-	43 (1,0)	-	43 (0,00)
Swiss-1 A	28 (0,06)	34 (0,07)	236 (0,50)	91 (0,19)	67 (0,14)	10 (0,02)	466 (0,03)
Swiss-1 B	24 (0,06)	41 (0,11)	88 (0,24)	96 (0,27)	91 (0,21)	13 (0,03)	353 (0,03)
Swiss-1 C	24 (0,07)	46 (0,14)	107 (0,33)	88 (0,27)	47 (0,14)	11 (0,03)	323 (0,02)
Swiss-1 D	10 (0,03)	83 (0,24)	88 (0,26)	77 (0,23)	54 (0,16)	21 (0,06)	333 (0,02)
Swiss-1 E	46 (0,11)	80 (0,20)	29 (0,07)	104 (0,26)	114 (0,29)	20 (0,05)	393 (0,03)

Swiss-2 A	-	-	81 (0,35)	61 (0,26)	66 (0,29)	18 (0,07)	226 (0,02)
Swiss-2 B	-	-	55 (0,21)	62 (0,24)	85 (0,33)	50 (0,19)	252 (0,02)
Swiss-2 C	-	-	72 (0,27)	75 (0,28)	99 (0,37)	20 (0,07)	266 (0,02)
Swiss-2 D	-	-	52 (0,25)	58 (0,28)	74 (0,36)	21 (0,10)	205 (0,02)
Swiss-2 E	-	-	47 (0,22)	58 (0,27)	75 (0,36)	28 (0,13)	208 (0,02)
Swiss (soma subgrupos)	132 (0,41)	284 (0,21)	855 (0,34)	770 (0,31)	772 (0,18)	212 (0,09)	3.025 (0,23)

*Valores exibidos na tabela estão apresentados em frequência absoluta e em entre parênteses frequência relativa.

Fonte: a autora

4.2.A série temporal indica maior produção de camundongos em 2018.

Através do conjunto de dados adquirido em cinco anos, geramos séries temporais a partir dos valores mensais. A decomposição dos componentes da série nos permite concluir que a produção de animais apresentou um gradativo aumento ao longo desse período, sendo sua produção superior em 2018 (Figura 4.4 e Tabela 4.5). A tendência (verde) clara e crescente demonstra um aumento na produção. O gráfico de sazonalidade representa o compilado de cada ano mês a mês nos 5 anos do estudo e é representada repetidamente em cada ano no gráfico (azul). Por último, o gráfico de ruídos, representados os dados que são aleatórios na série temporal (cinza).

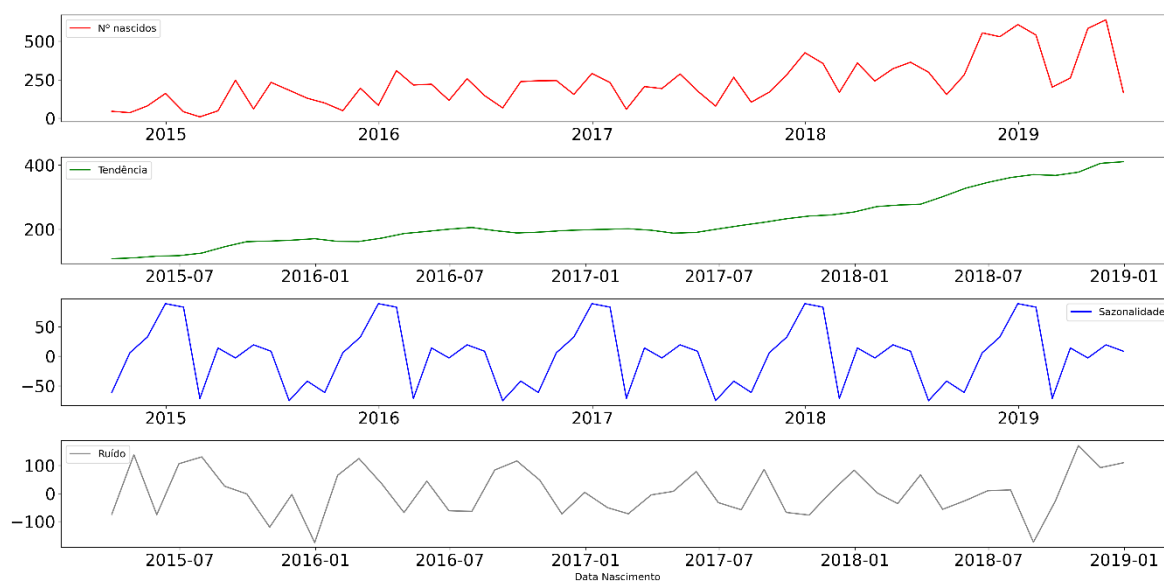


Figura 4.4 - Decomposição em série para os nascimentos de animais somando os meses de cada ano para todo o conjunto de dados. Representação gráfica do total de nascimentos na série temporal (vermelho) apresentando aumento gradativo, decomposição dos componentes de tendência (verde) acompanhando esse crescimento, sazonalidade (azul) e ruído (cinza).

Tabela 4.5- Número de nascimentos mensais distribuídos anualmente para todo o conjunto de dados.

ANO	2014	2015	2016	2017	2018	2019
Meses:						
Janeiro	-	42	310	232	356	541
Fevereiro	-	9	216	58	168	202
Março	-	48	221	206	359	262
Abril	-	247	116	193	242	584
Mai	-	60	257	288	322	640
Junho	-	233	157	177	364	166
Julho	-	182	67	79	299	-
Agosto	-	130	238	266	154	-
Setembro	45	99	244	104	281	-
Outubro	36	49	245	170	555	-
Novembro	80	195	155	283	530	-
Dezembro	161	84	291	426	609	-
TOTAL	322	1.378	2.507	2.482	4.239	2.395

*Sinal de (-) corresponde a valores não encontrados ou igual a zero.

Fonte: a autora

4.3. Produção do grupo isogênicas apresentam maior produção em 2018.

Segundo a análise da série temporal mensal do grupo isogênico, a produção foi superior no ano de 2018. Foi observado um crescimento da produção dessas linhagens entre o início de 2018 que se manteve até junho de 2019 (Figura 4.5). Os valores de nascimento para cada mês entre o período avaliado para o grupo isogênico estão descritos na Tabela 4.6. A tendência, novamente acompanha o crescimento de produção.

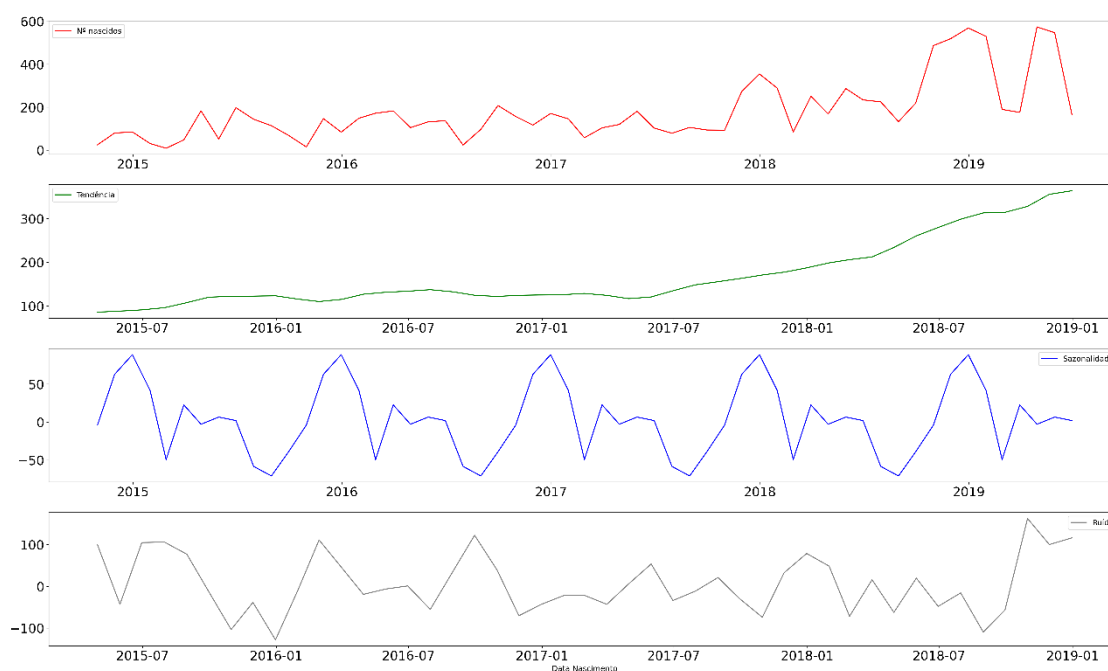


Figura 4.5 - Decomposição em série para os nascimentos de animais somando os meses de cada ano para o grupo Isogênico. Representação da decomposição em série entre os meses dentro do período de produção do grupo isogênico do biotério. O número de nascimentos (vermelho), tendência (verde) acompanhando esse crescimento, sazonalidade (azul), ruído (cinza).

Tabela 4.6- Número de nascimentos mensais distribuídos anualmente para o grupo Isogênico.

ANO	2014	2015	2016	2017	2018	2019
Meses:						
Janeiro	-	31	149	146	289	530
Fevereiro	-	9	179	58	85	190
Março	-	48	183	104	251	176
Abril	-	183	105	120	169	574
Mai	-	52	132	181	287	547
Junho	-	198	137	103	234	166
Julho	-	144	24	79	225	-
Agosto	-	114	97	106	133	-
Setembro	-	69	208	94	219	-
Outubro	25	15	157	93	487	-
Novembro	80	147	117	273	519	-
Dezembro	85	84	171	355	569	-
TOTAL	190	1.094	1.652	1.712	3.467	2.183

*Sinal de (-) corresponde a valores não encontrados ou igual a zero.

Fonte: a autora

4.4. Produção de linhagens heterogênicas apresentam maior produção no ano de 2016.

Na análise mensal, os dados do grupo heterogênico demonstram produção superior entre os anos de 2016 até 2018 (Figura 4.6), apresentando queda de produção no início de 2019, sendo esses valores também demonstrados na Tabela 4.7. A tendência se mostra crescente até a metade do ano de 2018, e demonstra uma queda de produção durante o ano de 2019, acompanhando os dados de produção.

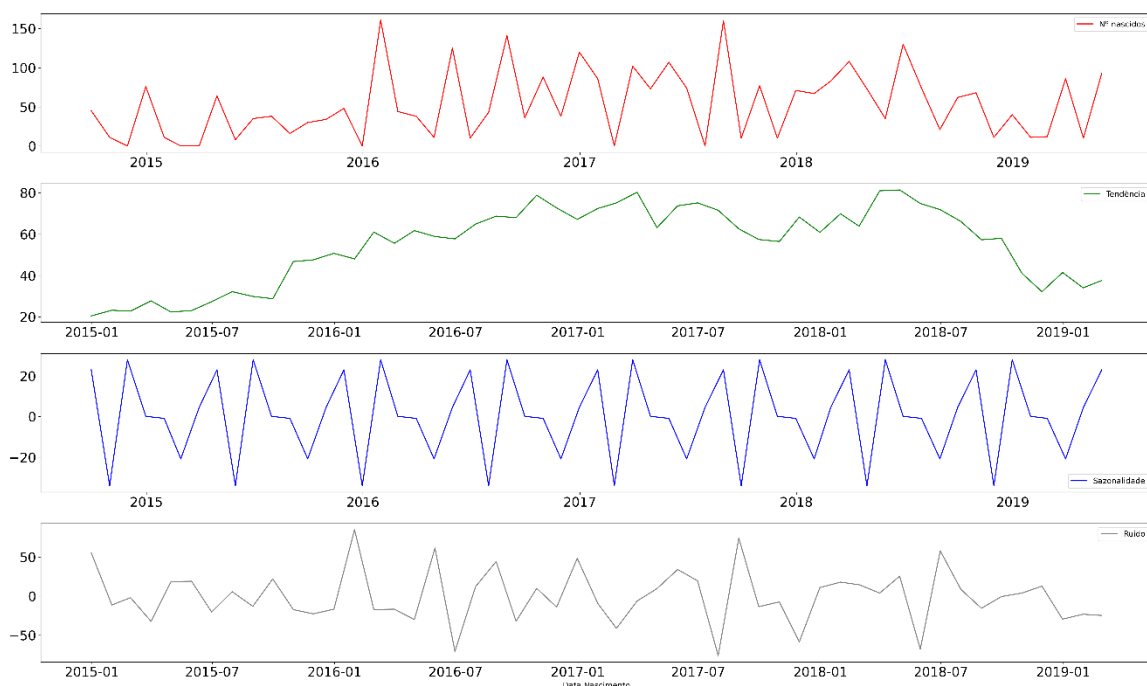


Figura 4.6 - Decomposição em série para os nascimentos de animais somando os meses de cada ano para o grupo Heterogênico. Representação da decomposição em série entre os meses dentro do período de produção do grupo Heterogênico do biotério. O número de animais (vermelho) apresentando crescimento até o ano de 2018, tendência (verde) acompanhando o crescimento até 2018 e posteriormente a queda em 2019, sazonalidade (azul), aleatoriedade (cinza).

Tabela 4.7- Número de nascimentos mensais distribuídos anualmente para o grupo Heterogênico.

ANO	2014	2015	2016	2017	2018	2019
Meses:						
Janeiro	-	11	161	86	67	11
Fevereiro	-	-	44	-	83	12
Março	-	-	38	102	108	86
Abril	-	64	11	73	73	10
Mai	-	8	125	107	35	93
Junho	-	35	10	74	130	-
Julho	-	38	43	-	74	-
Agosto	-	16	141	160	21	-
Setembro	-	30	36	10	62	-
Outubro	77	34	88	77	68	-
Novembro	10	48	38	10	11	-
Dezembro	71	-	120	71	40	-
TOTAL	132	284	855	770	772	212

*Sinal de (-) corresponde a valores não encontrados ou igual a zero. Fonte: a autora

4.5. O grupo Isogênico apresenta mediana de nascimentos inferior aos heterogênicos.

Ao calcular a mediana de nascimentos entre todos os registros (1.808 entradas), os dados retornaram 7,0 nascimentos/parto, já ao calcular a média, 7,67 nascimentos/parto. Ao buscar esses valores por grupo de linhagens de acordo com sua genética, foram obtidas medianas e médias diferentes para camundongos isogênicos e heterogêneos. Para o grupo isogênico (Figura 4.7 A), os dados apresentam média de 6,85 e mediana de 6,2 animais. Linhagens como DBA/1J, C57BL/6 e Balb/c apresentam valores acima da média e mediana, sugerindo que essas linhagens produzem maior número de nascimentos na colônia. Para o grupo heterogênico (Figura 4.7 B), a média e a mediana são muito próximas de aproximadamente 10,09 nascimentos.

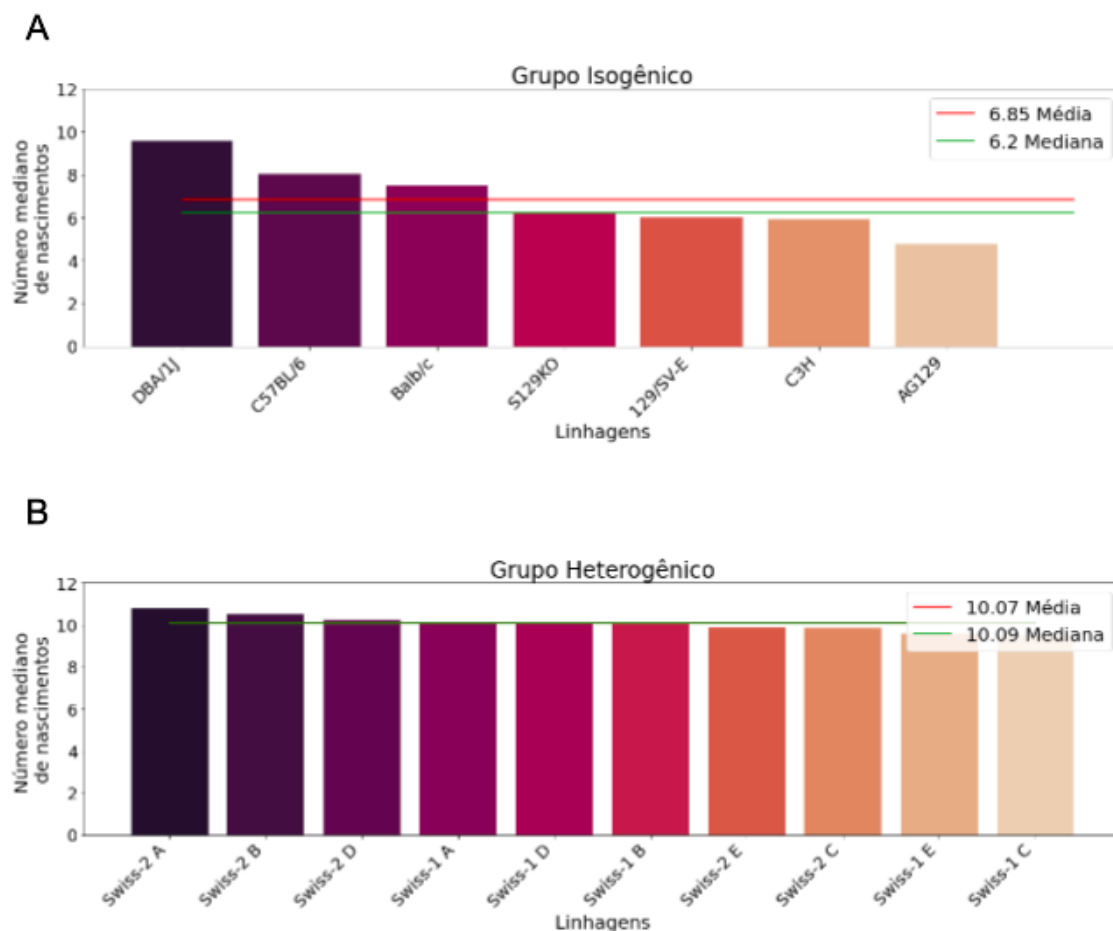


Figura 4.7 - Mediana de nascimentos segundo linhagem. (A) O grupo de linhagens isogênicas possui um valor menor de nascimentos em comparação ao grupo de linhagens heterogênicas. (B) A Linhagem Swiss (heterogênica) apresenta valores médios e medianos mais elevados do que as isogênicas.

4.6. A mediana do intervalo entre partos é em torno de 35 dias.

O intervalo entre partos (soma do acasalamento com o período gestacional da fêmea) foi analisado com base na mediana entre o dia em que o acasalamento começa até o nascimento dos camundongos (Figura 4.8). Maioria das linhagens pertencentes ao grupo isogênico, exceto DBA/1J, excederam a média (34,88/dias) e a mediana (35,0/dias) de aproximadamente 35 dias (acasalamento, gestação e nascimento).

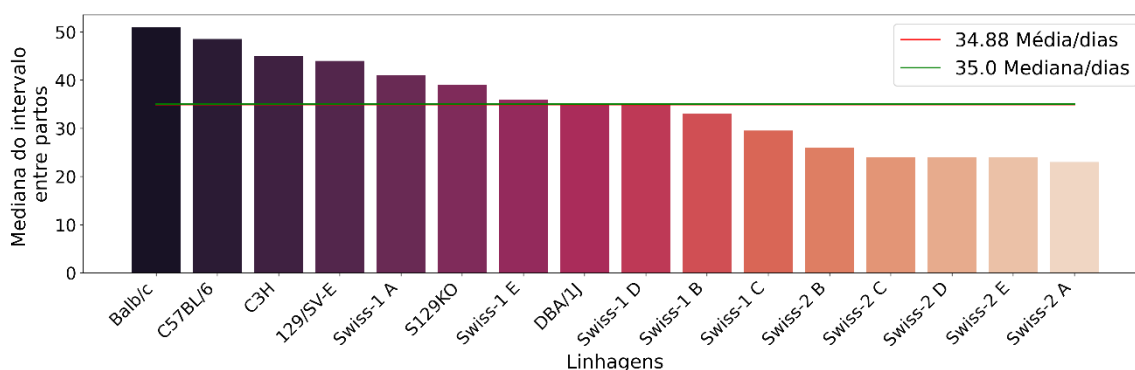


Figura 4.8 - Mediana do Intervalo entre partos para todas as linhagens. A maioria das linhagens do grupo isogênico apresentam valores superiores à média e mediana entre este período. As linhagens do grupo heterogênico requerem aproximadamente 22 a 30 dias para o nascimento.

4.7. A mediana entre o nascimento e o desmame dos camundongos é de 19 dias.

Ao atingir a idade do desmame, os filhotes dos camundongos são separados da mãe. A figura 4.9 apresenta a mediana entre o nascimento e desmame. Para todas as linhagens, a mediana desse período é de 19 dias, apenas AG129, C57BL/6 e 129/SV-E ultrapassam esse período. Ao analisar o gráfico, é possível observar que algumas das linhagens estão abaixo ou muito próximas da média, como o Swiss (grupo heterogênico), sendo esse intervalo ainda menor.

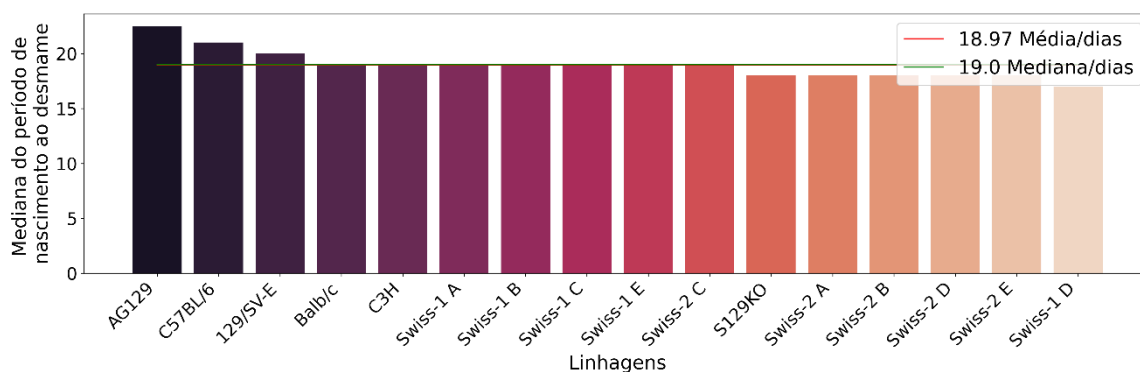


Figura 4.9 - Mediana do período entre o nascimento ao desmame para todas as linhagens. A maioria das linhagens se apresentam período de nascimento e o desmame em 19 dias.

4.8. As linhagens que são mais solicitadas entre usuários internos e externos do LACEA são C57BL/6 e Balb/c.

A partir do conjunto de dados referentes a solicitação de animais ao LACEA, foram analisados 403 registros com total de 5.776 animais solicitados. Na Tabela 4.8 estão descritos os registros de solicitação para cada ano.

Tabela 4.8- Número de entradas e de camundongos solicitados ao LACEA distribuídos anualmente.

ANO	Nº de Entradas		Nº de animais solicitados	
	Freq. Abs.	Freq. Rel.	Freq. Abs.	Freq. Rel.
2014	25	0,06	539	0,09
2015	70	0,17	766	0,13
2016	84	0,20	701	0,12
2017	78	0,19	910	0,15
2018	84	0,20	1.460	0,25
2019	62	0,15	1.400	0,24
Total	403	1,0	5.776	1,0

Fonte: a autora

Buscando entender o fluxo de fornecimentos dos camundongos pelo LACEA, os dados foram divididos entre solicitações internas, correspondendo os laboratórios do Instituto Carlos Chagas, e solicitações externas, de Instituições externas de ensino e pesquisa.

As linhagens mais frequentemente solicitadas por usuários internos são Balb/c (920, 59%), Swiss (365, 23%) e S129KO (202, 13%) (Figura 4.10 A e Tabela 4.9). Todas as demais linhagens estão apresentadas na Tabela 4.9. Podemos observar que a relação n° de animais/pedido é maior para as linhagens com background genético específico, S129KO (29,86) e AG129 (17,0) (Tabela 4.9). Dentre as solicitações externas, as linhagens mais frequentemente fornecidas durante o período de cinco anos analisado, são as linhagens C57BL/6 (1.780, 44%) com razão animais/pedidos de 37,08, Balb/c (1.392, 34%) razão 37,08 e Swiss (682, 17%) razão 21,31 (Figura 4.10 B e Tabela 4.9).

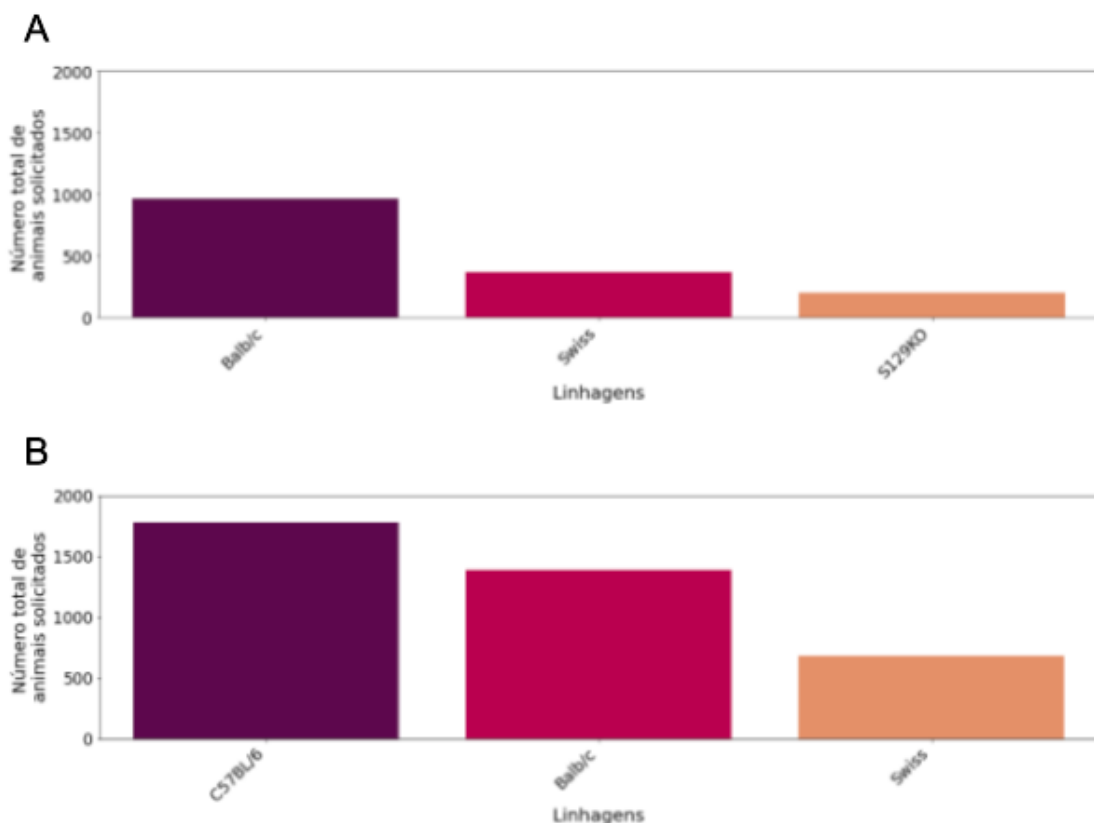


Figura 4.10 - Linhagens mais solicitadas ao LACEA entre usuários internos e externos. (A) Linhagens frequentemente solicitadas entre os pesquisadores internos do Instituto Carlos Chagas, sendo Balb/c a linhagem mais utilizada entre diferentes laboratórios. (B) Linhagens frequentemente solicitadas entre os usuários de instituições externas, são C57BL/6 e Balb/c.

Tabela 4.9- Número de entradas e soma de animais de cada linhagem solicitada entre os usuários internos e externos.

	Nº Entradas Freq. Abs.	Nº Entradas Freq. Rel.	Nº de animais solicitados Freq. Abs.	Nº de animais solicitados Freq. Rel.	Relação Animais/ Entradas
Usuários Internos:					
Balb/c	126	0,63	920	0,59	7,3
Swiss	61	0,3	365	0,23	5,98
S129KO	7	0,04	202	0,13	29,86
C57BL/6	3	0,02	25	0,02	8,33
AG129	2	0,01	34	0,02	17,0
DBA/1J	1	0,001	10	0,01	10,0
TOTAL	200	1,0	1.556	1,0	7,78
Usuários Externos					
C57BL/6	48	0,27	1,780	0,44	37,08
Balb/c	78	0,43	1,392	0,34	37,08
Swiss	32	0,18	682	0,17	21,31
129/SV-E	6	0,03	88	0,02	14,67
S129KO	14	0,08	85	0,02	6,07
C3H	2	0,01	14	0,0	7,0
TOTAL	180	1,0	4.041	1,0	22,45

Fonte: a autora

Entre as informações das instituições externas, buscamos quais delas apresentaram maior número de animais solicitados ao LACEA. As instituições que mais fizeram solicitações foram UFPR (Universidade Federal do Paraná) com total de 1.647 camundongos solicitados e UEL (Universidade de Londrina) com 1.195 (Figura 4.11). As instituições, com o número de animais solicitados, estão descritas na Tabela 4.10. As instituições com a maior razão número de animais/pedido foram o IBMP (66,0) seguido da UEL (56,9). Dentre as instituições presentes, podemos observar

registros como "ICC", "Biotério" e "Externo", sugerindo a presença de mais um erro de registro.

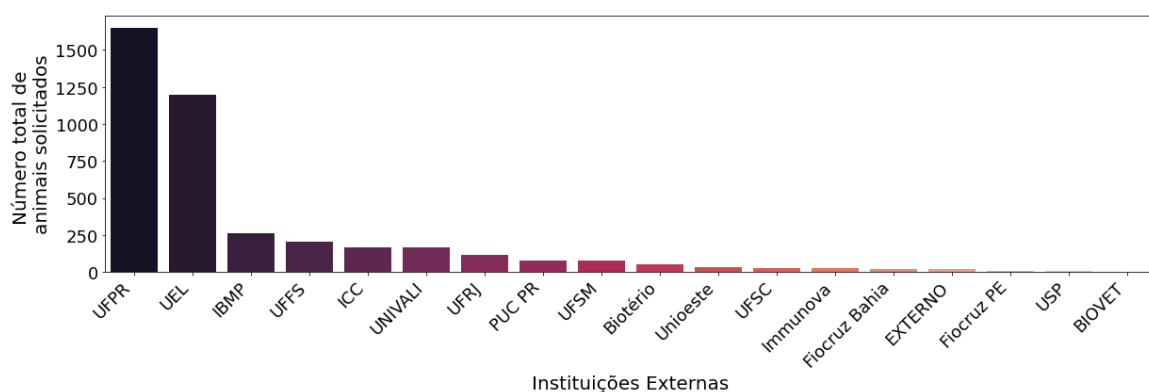


Figura 4.11- Instituições externas que mais solicitaram animais ao LACEA no período de cinco anos. Entre as instituições externas que solicitam camundongos pelo LACEA, a mais frequente é a UFPR e em seguida UEL.

Tabela 4.10- Número de entradas e a soma de animais solicitados ao LACEA por instituições externas.

Instituições Externas	Nº Entradas	Nº Entradas	Nº de animais	Nº de animais	Relação Animais/Entradas
	Freq. Abs.	Freq. Rel.	solí. Freq. Abs	solí. Freq. Rel.	
UFPR	68	0,37	1.647	0,4	24,22
UEL	21	0,11	1.195	0,29	56,9
IBMP	4	0,02	264	0,06	66,0
UFFS	4	0,02	208	0,05	52,0
ICC	40	0,22	167	0,04	4,18
UNIVALI	5	0,03	164	0,04	32,8
UFRJ	6	0,03	115	0,03	19,17
PUC PR	5	00,3	80	0,02	16,0
UFSM	2	0,01	78	0,02	39,0
Biotério	5	0,03	50	0,01	10,0
UNIOESTE	1	0,01	36	0,01	36,0
UFSC	8	0,04	29	0,01	3,62
Immunova	1	0,01	27	0,01	27,0

Fiocruz	1	0,01	20	0,0	20,0
Bahia					
Externo	8	0,04	20	0,0	10,0
Fiocruz PE	1	0,01	10	0,0	10,0
USP	2	0,01	3	0,0	3,0
Biovet	1	0,01	3	0,0	3,0
TOTAL	183	1,0	4.121	1,0	22,52

Fonte: a autora

4.9. O principal motivo de descarte de camundongos é por Descarte zootécnico.

Entre os dados de descarte animal, foi encontrado um total de 8,754 animais descartados entre 2.286 registros (entradas). O número de animais descartados a cada ano está descrito na Tabela 4.11

Tabela 4.11- Número de entradas e de camundongos descartados distribuídos anualmente.

ANO	Nº de Entradas	Nº de Entradas	Nº de animais	Nº de animais
	Freq. Abs.	Freq. Rel.	descartados	descartados
			Freq. Abs.	Freq. Rel.
2014	42	0,01	233	0,02
2015	304	0,13	996	0,11
2016	482	0,21	1.607	0,18
2017	452	0,19	1.678	0,19
2018	665	0,29	2.197	0,25
2019	341	0,14	1.170	0,13
Total	2.286	1,0	8.754	1,0

Fonte: a autora

Entre todos os motivos de descarte de camundongos, o motivo Descarte Zootécnico foi o que apresentou maior número de animais (2.986 camundongos, 46%) descartados no LACEA, com uma razão de 3,1 animais descartados para cada

registro realizado, sendo um dos motivos mais comuns de descarte entre os biotérios (Figura 4.12 A). Esse motivo geralmente utilizado para registrar animais que foram descartados por brigas ou caso tenham sido produzidos em excesso. Na Tabela 4.12 os dados de descarte são apresentados segundo motivo pelo qual o camundongo foi descartado, sendo um deles representado por “Outros”, onde foi somado todos os registros que não eram descartes reais e sim uma transferência de caixa, novamente apresentando erro de registro.

Tabela 4.12- Número de entradas e soma de animais segundo os principais motivos de descarte animal para todo conjunto de dados.

	Nº Entradas Freq. Abs.	Nº Entradas Freq. Rel.	Nº de animais descartados Freq. Abs.	Nº de animais descartados Freq. Rel.	Relação Animais/ Entradas
Motivos de Descarte:					
Descarte Zootécnico	962	0,55	2.986	0,46	3,1
Fim Experimento	316	0,13	1.321	0,15	4,18
Encontrado Morto	163	0,07	859	0,09	5,27
Canibalismo	73	0,03	598	0,06	8,19
Velhos	110	0,04	358	0,04	3,25
Fornecido Zoológico	71	0,03	185	0,02	2,61
Animais Doentes	36	0,01	128	0,01	3,56
Ivermectina	15	0,01	50	0,005	3,33
Monitoramento Sanitário	17	0,006	42	0,004	2,47
Outros	523	0,22	2.227	0,25	4,2
TOTAL	2.286	1,0	8.754	1,0	3,7

Fonte: a autora

Entre o conjunto de dados apenas para o grupo isogênico, foram encontrados o total de 5.814 animais em 2.072 entradas, razão de 2,8 animais descartados/registro. Para os dados do grupo heterogênico, foram identificados o total de 2.940 camundongos descartados entre 571 entradas, razão de 5,14 animais descartados/registro. Em ambos os grupos, novamente o Descarte Zootécnico, 2.061 camundongos descartados pelo grupo Isogênico (45%) (Tabela 4.13) e 925 camundongos descartados pelo grupo heterogênico (47%) (Tabela 4.14), se apresentou como principal motivo de descarte entre os camundongos (Figura 4.12 B-C). Os motivos de descarte nos diferentes grupos são apresentados na Tabela 4.13-14.

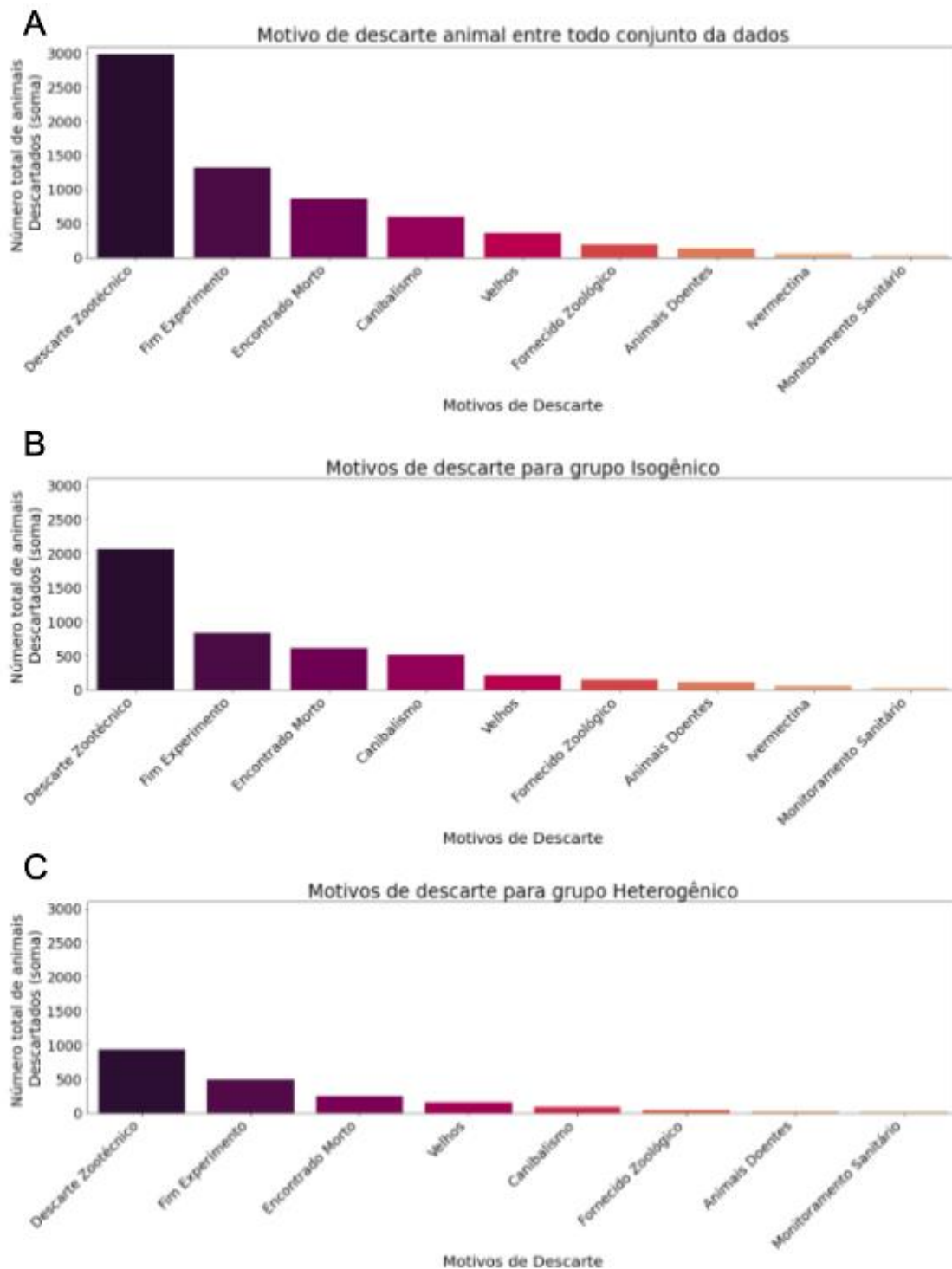


Figura 4.12- Principais motivos de descarte animal registrados pelo LACEA. (A) Total de animais que foram descartados para cada motivo de descarte. (B) Total de animais descartados por motivos de descarte para o grupo isogênico. (C) Total de animais descartados por motivos de descarte para o grupo heterogênico.

Tabela 4.13- Número de entradas e soma de animais segundo os principais motivos de descarte animal para o grupo Isogênico.

	Nº Entradas Freq. Abs.	Nº Entradas Freq. Rel.	Nº de animais descartados Freq. Abs.	Nº de animais descartados Freq. Rel.	Relação Animais/ Entradas
Motivos de Descarte:					
Descarte Zootécnico	694	0,55	2.061	0,45	2,97
Fim Experimento	204	0,16	833	0,10	4,08
Encontrado Morto	128	0,1	615	0,14	4,8
Canibalismo	64	0,05	510	0,11	7,97
Velhos	60	0,05	208	0,05	3,47
Fornecido Zoológico	56	0,04	142	0,03	2,54
Animais Doentes	30	0,02	106	0,02	3,53
Ivermectina	15	0,01	50	0,01	3,33
Monitoramento Sanitário	12	0,01	25	0,01	2,08
Outros	239	0,3	1.264	0,21	1,68
TOTAL	1.502	1,0	5.814	1,0	2,80

Fonte: a autora

Tabela 4.14- Número de entradas e soma de animais segundo os principais motivos de descarte animal para o grupo Heterogênico.

	Nº Entradas Freq. Abs.	Nº Entradas Freq. Rel.	Nº de animais descartados Freq. Abs.	Nº de animais descartados Freq. Rel.	Relação Animais/ Entradas
Motivos de Descarte:					
Descarte Zootécnico	268	0,54	925	0,47	3,45
Fim Experimento	112	0,22	488	0,25	4,36
Encontrado Morto	35	0,07	244	0,12	6,97
Canibalismo	50	0,1	150	0,08	3,0
Velhos	9	0,02	88	0,04	9,78
Fornecido Zoológico	9	0,02	88	0,04	9,78
Animais Doentes	6	0,01	22	0,01	3,67
Monitoramento Sanitário	5	0,01	17	0,01	3,4
Outros	71	0,12	963	0,32	13,56
TOTAL	571	1,0	2.940	1,0	5,14

Fonte: a autora

Depois da análise de descarte entre os grupos de animais isogênicos e heterogênicos, foi analisado os motivos de descarte da linhagem C57BL/6 uma vez que esta é a linhagem mais produzida no LACEA. Foi observado um total de 1.451 camundongos descartados entre 330 entradas, razão de 4,39 descartes/registro. O principal motivo de descarte é o Descarte Zootécnico (455 camundongos, 39%), em seguida Encontrado Morto (323 camundongos, 27%) e em terceiro lugar por Canibalismo (269 camundongos, 23%), por se tratar de uma linhagem agressiva, é esperado esses motivos de descarte com mais registros (Figura 4.13 A). Os demais motivos de descarte para essa linhagem estão apresentados na Tabela 4.15.

Ao analisar os motivos entre machos (Figura 4.13 B) e fêmeas (Figura 4.13 C), em ambas as condições, ao observar as nove principais razões de descarte, o maior motivo foi o Descarte Zootécnico (174, 68% camundongos machos descartados e 281, 84% camundongos fêmeas descartadas). Entre os dados dos camundongos machos, não há registros sobre fim de experimento, sugerindo que animais machos dessa linhagem, não são significativamente usados em experimentação. Já entre os registros de camundongos fêmeas, apresentam como segundo principal motivo de descarte o Fim de Experimentação, corroborando a hipótese. Os dados de descarte de camundongos machos e fêmeas estão apresentados na Tabela 4.16.

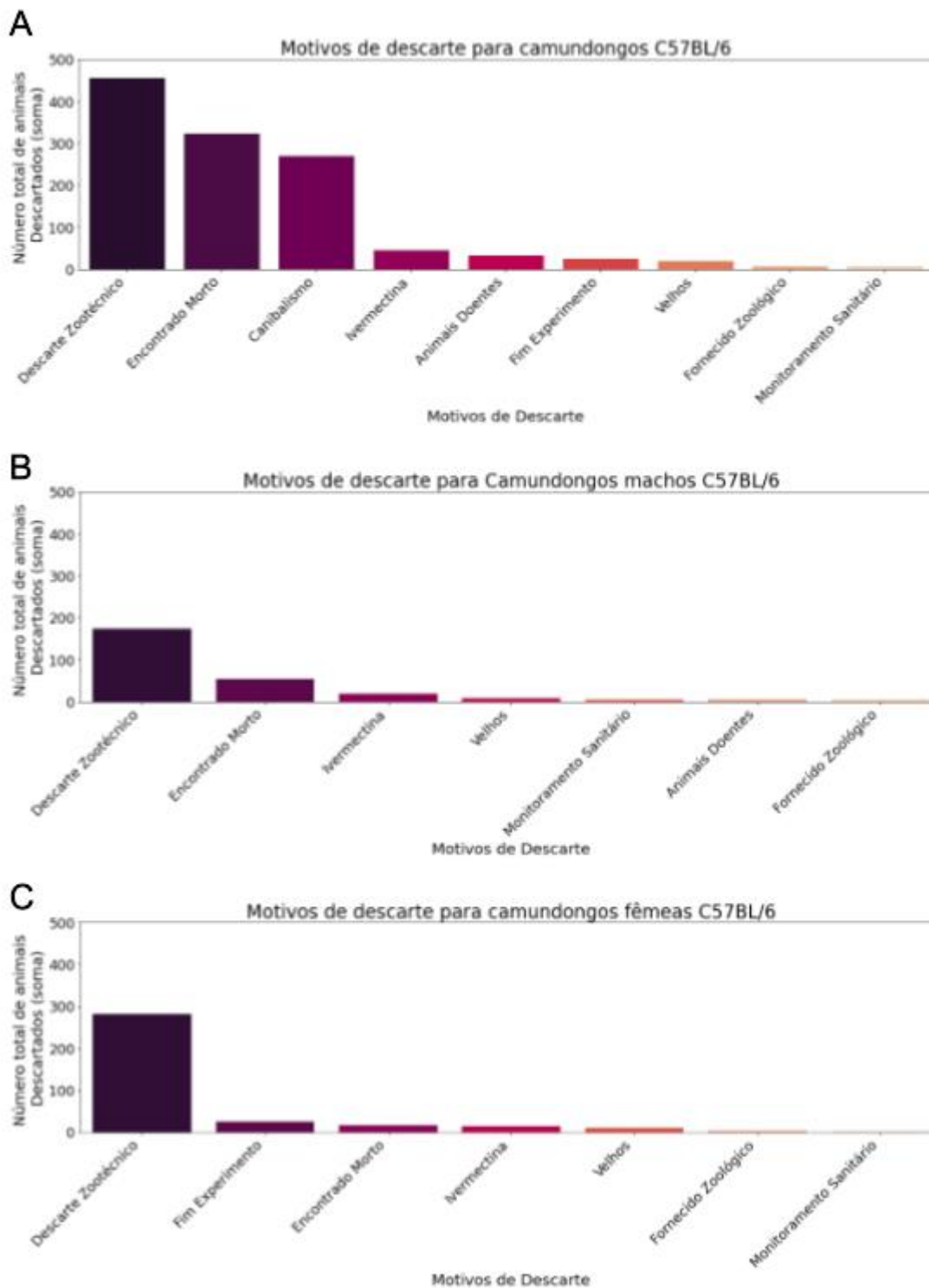


Figura 4.13- Principais motivos de descarte da linhagem C57BL/6. (A) Representação da soma de todos os motivos de descarte registrados para a linhagem C67BL/6 (B) Soma do descarte de camundongos machos C57BL/6. (C) Soma do descarte de camundongos C57BL/6 fêmeas.

Tabela 4.15 - Número de entradas e soma de animais segundo os principais motivos de descarte animal para a linhagem C57BL/6.

	Nº Entradas Freq. Abs.	Nº Entradas Freq. Rel.	Nº de animais descartados Freq. Abs.	Nº de animais descartados Freq. Rel.	Relação Animais/ Entradas
Motivos de Descarte:					
Descarte Zootécnico	142	0,43	455	0,31	3,2
Encontrado Morto	63	0,19	323	0,22	5,13
Canibalismo	35	0,11	269	0,19	7,69
Ivermectina Animais Doentes	12	0,04	45	0,03	3,75
	6	0,02	32	0,02	5,33
Fim Experimento Velhos	2	0,01	25	0,02	12,5
	7	0,02	19	0,01	2,71
Fornecido Zoológico	3	0,01	7	0,01	2,33
Monitoramento Sanitário	3	0,01	6	0,00	2,0
Outros	57	0,17	270	0,19	4,73
TOTAL	330	1,0	1.451	1,0	4,39

Fonte: a autora

Tabela 4.16 - Número de entradas e soma de animais segundo os principais motivos de descarte animal para machos e fêmeas da linhagem C57BL/6.

	Nº Entradas Freq. Abs.	Nº Entradas Freq. Rel.	Nº de animais descartados Freq. Abs.	Nº de animais descartados Freq. Rel.	Relação Animais/ Entradas
Motivos de Descarte para machos de C57BL/6:					
Descarte Zootécnico	57	0,66	174	0,68	3,05
Encontrado Morto	15	0,17	54	0,21	3,6
Ivermectina	6	0,07	20	0,08	3,33
Velhos	4	0,05	9	0,04	2,25
TOTAL	82	1,0	257	1,0	12,23
Motivos de Descarte para fêmeas de C57BL/6:					
Descarte Zootécnico	85	0,89	281	0,81	3,31
Fim Experimento	2	0,02	25	0,07	12,5
Encontrado Morto	5	0,05	15	0,04	3,0
Ivermectina	4	0,04	14	0,04	3,5
TOTAL	96	1,0	355	1,0	22,31

Fonte: a autora

Entre as movimentações do descarte animal, foram observadas uma série de registros anômalos. Além dos registros caracterizados como algum tipo de descarte, também eram registradas transferências, ou registrados como uma maneira de "rastrear" algum procedimento de manejo da colônia que não pudesse mais ser modificado nos processos próprios do BioterC.

Ao compilar esses dados, entre os motivos mais comuns foi encontrado o "Erro de Cadastro" apresentando 97 registros (543 camundongos, 24%), "Para Acasalamento", em que foi observado 53 registros (94 camundongos, 4%), sugerindo que sejam dados de transferências de casais para montagem de acasalamento e "UEL" apresentando 40 registros (143 camundongos, 6%), sugerindo que esses registros sejam na verdade dados de camundongos fornecidos a essa instituição. Além dos motivos identificados como um erro de registro, 10% dos registros que não se encaixavam a nenhum motivo de descarte, ou entre os motivos de manejo da colônia apresentados na Tabela 4.17, também foram considerados como um erro de cadastro, os quais perfazem 45% dos registros anômalos.

Tabela 4.17- Registros na sessão de descarte do BioterC que foram representados como "Outros" por não serem motivos de descarte animal.

	Nº Entradas Freq. Abs.	Nº Entradas Freq. Rel.	Nº de animais Freq. Abs.	Nº de animais Freq. Rel.	Relação Animais/ Entradas
Motivos que não representam descarte:					
Erro de Cadastro	97	0,33	543	0,24	5,59
Para Acasalamento	53	0,18	94	0,04	1,77
UEL	40	0,13	143	0,06	3,57
Mudança	25	0,08	68	0,03	2,72
Transferência Cx	23	0,08	78	0,03	3,39

Troca de Macho	21	0,07	53	0,02	2,52
Excesso de Filhotes	14	0,04	180	0,08	12,82
Ex Lorena	13	0,04	47	0,02	3,61
10% dos motivos que não representa descarte/erro	236	0,00	1.021	0,45	1.201
TOTAL	523	1,0	2.227	1,0	0,79

Fonte: a autora

4.10. As análises dos dados do BioterC direcionaram modificações gerenciais no LACEA.

Após apresentar os resultados obtidos através dos dados do BioterC para a equipe do LACEA, foram realizadas mudanças para buscar melhorias do gerenciamento da colônia de animais deste biotério. Com a apresentação dos *outliers* presentes nos dados de nascimentos, houve mudanças de comportamento entre a equipe técnica do LACEA quanto ao modo de realização dos registros das informações da colônia. Na nova versão desse *software*, BioterCS, foi reforçada a necessidade de preenchimento de acordo com o número de nascimentos, e incluído ainda na raiz do programa a possibilidade de edição dos dados, caso ocorra um erro ao realizar o registro.

Para reduzir o período de intervalo entre partos de 35 dias, os bioteristas passaram a ciclar todas as fêmeas em estado de anestro (ausência de cio). Para isso, as fêmeas, 2-3 dias antes da montagem dos acasalamentos, são expostas aos feromônios dos machos ao receber em suas caixas a maravalha suja oriunda das caixas dos machos.

Outra mudança relacionada ao acasalamento dos camundongos, foi montar os casais com animais mais experientes. Os acasalamentos antes eram formados por casais de 45 dias de idade (jovens) e utilizados por, no máximo, duas gestações consecutivas e então descartados. Hoje, buscando reduzir o intervalo entre partos, a equipe de bioteristas passou a utilizar animais com 60 dias de idade, e chegar a retirar

de 4-6 ninhadas de cada casal formado. Essa mudança, além de reduzir o descarte prematuro dos animais utilizados, também possibilita a formação de uma colônia com animais mais experientes. Após a identificação de linhagens com ausência de solicitações, outra medida de melhoria implantada pela equipe do LACEA foi a de eliminar da colônia estas linhagens, como por exemplo, a linhagem C3H.

Já as linhagens com pouco solicitações, por sua vez, estão sendo mantidas com apenas um único casal para fins de renovação da matriz. Antes, essas colônias eram mantidas com 3-5 casais, concomitantes, para cada linhagem.

Outra nova medida visando a redução de descartes implantada no LACEA, foi atender somente pedidos formalmente realizados através do *software*, reduzindo a prática de manutenção de animais de prateleira, comumente usados nos biotérios brasileiros, que tem o objetivo de atender de forma contínua mesmo aqueles pedidos sem previsão.

Os pedidos de instituições externas, revelaram o fornecimento de animais para laboratórios internos que foram registrados como sendo pedidos externos. O diálogo junto com a equipe revelou que existia a prática pontual de fornecer animais internamente para laboratórios que não possuíam CEUA vigente ou ainda em trâmite de análise. A presente situação foi totalmente resolvida após a coordenação do biotério fornecer pedidos apenas para laboratórios que possuem CEUA válida, cadastrada no BioterC.

Além disso, ao apresentar os principais motivos de descarte aos bioteristas do LACEA, novas atitudes vêm sendo continuamente implementadas para reduzir o número de animais descartados, a saber:

1) Aumento do período de permanência dos casais na colônia matriz e de produção, com a separação dos casais. Para isso, o biotério passou a usar “haréns” temporários, onde os machos e fêmeas são agrupados somente quando é necessária a produção de animais para atender os pedidos. Após constatada a gestação, cerca de 12 dias de crescimento fetal, os machos são separados para evitar a cobertura do cio pós-parto.

2) Com relação aos animais excedentes, isto é, aqueles que sobram de um pedido realizado, a equipe passou a fazer blocos sincronizados de clientes de diferentes instituições. Para tanto, a equipe tornou-se menos passiva e mais ativa na parte de organização dos cronogramas de pedidos, sugerindo datas alternativas de fornecimento para os usuários.

3) Ainda, com o objetivo de aumentar as vendas, mesmo de linhagens com pouco uso, o biotério passou a utilizar estratégias de marketing virtual, com a produção de material de conteúdo do biotério, vídeos do setor e o estabelecimento de um *WhatsApp* comercial para se comunicar mais rapidamente com os clientes. Caso não seja encontrado compradores para estes animais, os neonatos passaram a ser descartados. Isso levou a redução da manutenção de animais sem destino na sala, reduzindo mão de obra, bem com os gastos de recursos públicos tais como ração, maravalha, desinfetantes e energia elétrica, descarte de resíduos e utilização do maquinário.

4) Entre os motivos de descarte para a linhagem C57BL/6, foram observados índices consideráveis de canibalismo. Na tentativa de contornar essa situação, a equipe do biotério elevou para 100% o enriquecimento ambiental nas caixas. É importante que se ressalte que este índice, previamente ao trabalho em tela, antes não alcançava 20% da colônia. Para que fosse possível alcançar e manter o índice de enriquecimento ambiental de 100% das caixas do biotério, a equipe de bioteristas, junto com a Assessoria de Comunicação do Instituto, tem feito campanhas em fluxo contínuo dentro do ICC, junto aos usuários do biotério, para que se sensibilizem em trazer materiais de enriquecimento ambiental para os animais. Outra medida gerencial tomada, ainda nessa linha, foi o ajuste e organização de grupos de comunicação com a infraestrutura do ICC. Como resultado, a melhoria da comunicação interna tem agilizado os reparos no ar-condicionado e, conseqüentemente, minimizado os eventos de alteração de temperatura, pressão e umidades dentro das salas dos animais, fatores estes que tem contribuição direta com eventos de canibalismo.

5. Discussão

Dentre as análises do conjunto de dados, obtemos resultados relevantes da criação animal dentro de um biotério. Entre nossos achados, padrões de criação foram confirmados, tais como mediana de nascimento entre linhagens de constituição genética isogênicas e heterogênicas. Além de identificar padrão de criação, também foi possível obter os resultados de descarte animal, que apresentam valores discrepantes em comparação à produção e solicitação.

Os resultados obtidos através da Mineração de Dados, utilizando como metodologia conceitos de Ciência de Dados, facilitou a triagem e análise dos dados, chegando a resultados de uma maneira ágil e confiável. Todas as rotinas de análise foram desenvolvidas durante o trabalho, não sendo necessário recorrer a soluções pagas, sendo apenas necessário conhecimento prévio de linguagem de programação.

O primeiro objetivo do presente trabalho foi identificar no conjunto de dados o número de animais que foram produzidos no LACEA no período de 2014 a 2019. A análise inicial mostrou a presença de registros discrepantes, o que foi comprovado quando utilizamos gráficos de histograma e *boxplot*, onde foi possível encontrar *outliers*, sendo necessário ajustar esses dados dentro de uma distribuição aproximada de número de nascimentos por evento de nascimento de acordo com o que é observado na rotina pelos bioteristas.

O total de camundongos produzidos, após retirada de valores discrepantes, foi de 13.323 camundongos entre 1.736 entradas. Não foi possível realizar um comparativo a partir desse dado com outro biotério que apresente produção de camundongos próximo ao que foi avaliado, uma vez que esses dados não estão disponíveis para confronto de produção animal, limitando nossas investigações.

De acordo com Chatfield (2000), uma Série Temporal é um conjunto de observações medidas sequencialmente ao longo do tempo. Assim, buscamos apresentar séries temporais com base em dados mensais, a partir do conjunto de dados dentro de um período de cinco anos. Ao analisar os dados como um todo, foi possível observar que o período do fim de 2018 até 2019, apresentou registros superiores de produção, ou seja, tendência crescente. Ao traçar a mesma série para o grupo de animais isogênicos, os valores se mantêm, porém, para o grupo

heterogênico, observamos queda de produção em 2018, sendo esses valores refletidos na tendência.

Ao calcular a mediana entre valores de nascimentos, obtivemos valores diferentes entre os dois grupos analisados, a saber, isogênicos (acasalamento irmão x irmão), são um grupo de linhagens geneticamente definidas com alta consanguinidade (98,6% de endogamia). Trata-se de camundongos com menor desempenho reprodutivo se comparados com os animais do grupo heterogênico. Já o grupo heterogênico (acasalamento randômico), são geneticamente variáveis e mantidos de maneira que se evite ao máximo a consanguinidade e se preserve a variabilidade genética.

Nossos resultados corroboram com Benavides e colaboradores (2003) que afirma que, os camundongos geralmente parem de 4-8 filhotes em linhagens isogênicas, e no caso de alguns grupos de camundongos (heterogênicos), o número de filhotes pode variar de 10 a 14, com um registro de 32 filhotes para uma fêmea do grupo de camundongos suíços (Swiss). De acordo com Santos e colaboradores (2002), a média de filhotes/parto é de 8-10 em linhagens heterogênicas (podendo ser chamada de *outbred*) e em torno de 5 filhotes/parto em linhagens isogênicas (podendo ser chamada de *inbred*). Apesar das linhagens isogênicas apresentarem menor prole, são modelos com menor variabilidade genética, podendo reduzir o número de animais que forem utilizados, devendo apenas se questionar qual o melhor modelo para se usar. Como propõe Braga e colaboradores (2017) o conhecimento da genética desses animais é fundamental para que se construam controles reprodutivos capazes de evitar desperdícios financeiros, uso inadequado do espaço do biotério e sobretudo o descarte excessivo de excedentes.

A mediana do intervalo entre partos (IEP) é de 35 dias. O IEP corresponde o período da retomada do ciclo estral após desmame do parto da fêmea somado ao período de gestação da camundonga. A fêmea entra em ciclo estral (cio), que tem duração de 4-5 dias. Quando expostas aos machos ou a seus feromônios, começam a ciclar em 48 horas, quando constatada a prenhez, o período de gestação vai de 19 a 21 dias (Santos, 2002). O período em que os animais permanecem em acasalamento é de 16 dias, sendo 11 dias a mais do que está descrito na literatura. Para redução desse intervalo, os bioteristas estão trabalhando com acasalamento programado, o com o objetivo de sincronizar os nascimentos e os desmames, para

obter fornecimento programado, produzindo o mínimo de animais excedentes. Para o grupo isogênico, observamos mediana de 48,80 dias, sendo 12 dias a mais do que os heterogênicos (36,51 dias).

Seguindo com os processos de criação animal, foi calculada a mediana entre o período do nascimento e o desmame dos camundongos. Esse período é determinado como idade de desmame dos filhotes, que geralmente ocorre em 19 dias após o nascimento, sendo que em linhagens do grupo isogênico o desmame pode ocorrer em até 28 dias (Santos, 2002). No momento do desmame, os animais são sexados e separados. Guénet e colaboradores (2015) explica que a amamentação dura entre 19 e 21 dias, onde os camundongos estariam prontos para alimentação de sólidos e preparados para serem sexados. Mesmo que dentro do período descrito, a média e mediana de 19 dias encontrados nos dados, indicam precocidade da sexagem.

Kikusi e colaboradores (2005) demonstraram que o comportamento de filhotes e mães separados prematuramente pode ser alterado, com maior atividade dos camundongos e menor repouso durante o período da pós-sexagem, causando ansiedade nos filhotes da colônia. Uma vez que a produção animal foi caracterizada, buscou-se analisar os registros de solicitações de animais ao LACEA através do BioterC.

O número de animais solicitados por usuários externos é três vezes (4.041 camundongos) superior ao solicitado por usuários internos (1.556 camundongos), porém, ocorre em menor frequência (usuários externos 180 entradas e usuários internos 200 entradas). As linhagens mais solicitadas entre os usuários do ICC são Balb/c, Swiss e S129KO, sendo a linhagem Balb/c solicitada pela maioria dos laboratórios. Já a linhagem Swiss deixou de ser solicitada no ano de 2018, onde pode ser observado na série temporal a queda de produção para esse grupo. A partir desse dado, levantou-se a hipótese de que essa linhagem tenha sido substituída por Balb/c na experimentação, pelo propósito dos pesquisadores buscarem por redução do uso de animais e refinamento em seus experimentos ao utilizar linhagens isogênicas, garantindo menor variabilidade genética em seus resultados.

Entre as instituições externas que mais solicitam camundongos ao LACEA estão a UFPR (Universidade Federal do Paraná), a UEL (Universidade Estadual de Londrina) e o IBMP (Instituto de Biologia Molecular do Paraná). Foi realizada uma pesquisa interna no biotério, gerando um relatório de *“Criação e Experimentação*

Animal Fiocruz PR”, buscando o perfil de clientes externos que solicitam animais para entender qual o principal motivo de buscarem camundongos do LACEA. Segundo esse relatório, 90% das solicitações são de instituições de pesquisa externa e o motivo principal para solicitação é devido à alta disponibilidade de animais no biotério, além de fornecerem animais de qualidade (Kracker et al, 2019).

Com o presente trabalho, foram aplicadas modificações gerenciais no LACEA em relação ao fornecimento, como por exemplo, o fortalecimento da prática de fornecer animais dentro da CEUA, com validade não expirada e com quantidade de animais disponíveis. Essa mudança de postura impulsionou os pesquisadores a passarem a redigir os pedidos de licença CEUA com antecedência e aproximou os usuários da equipe técnica do LACEA que participa ativamente na revisão desses pedidos com o objetivo de garantir a rápida aprovação das licenças.

O descarte é a última das etapas da criação animal. De acordo com as diretrizes para eutanásia estabelecida pelo CONCEA, os critérios frequentemente adotados para recomendação de eutanásia de uma forma individualizada são: animais feridos, com impossibilidade de tratamento, animais com doenças terminais em sofrimento intenso e animais idosos na falta de recursos para atender às suas necessidades. Porém, outras situações que indicam a indução da morte podem ocorrer, como, por exemplo, quando os animais forem submetidos a atividades de ensino ou de pesquisa científica. De acordo com o CONCEA (2013) a eutanásia exige considerações morais e éticas para que a prática seja realizada de forma humanitária.

O motivo de descarte mais frequente no biotério é o de Descarte Zootécnico, que geralmente está relacionado com camundongos descartados por briga, animais que estão velhos, que não podem mais ser usados para reprodução ou são criados em excesso para a manutenção da colônia. Cardoso e colaboradores (2002) afirmam que, considerando a condição física de um biotério, é desejável uma maior produtividade com menor custo possível, sendo o descarte de animais indesejável ao plantel. Em todo conjunto de dados de descarte do BioterC, o Descarte Zootécnico apresentou 2.986 animais (46% dos camundongos) eutanasiados, representando mais que o dobro de descartes quando comparado ao motivo Fim de Experimento que apresentou 1.321 animais (15% dos camundongos) eutanasiados.

Ao analisar os dados de descarte segundo animais do grupo isogênico e heterogênico, observou-se que os grupos apresentam Descarte Zootécnico (isogênico um total de 2.061 camundongos descartados e heterogênico um total de

925 camundongos descartados) superior em relação aos animais que foram para experimentação (isogênico 833 camundongos e heterogênicos 488).

Entre os dados de descarte da linhagem C57BL/6, os principais motivos de descarte registrados foram de Descarte Animal, Encontrado Morto e Canibalismo, sendo motivos de descarte comum entre essa linhagem, pois de acordo com os bioteristas, esses animais apresentam comportamento dominante e agressivo. Brown e colaboradores (1999), afirmam que entre as fêmeas de C57BL/6 podem apresentar distúrbios de comportamento materno, sendo mais propensa a apresentar infanticídio comparadas as outras linhagens.

Com novas medidas de montagem de acasalamento, essa prática tem minimizado os canibalismos entre as fêmeas, aumentando o cuidado maternal. Para evitar casos de canibalismo entre as linhagens, outra medida tomada foi nos alojamentos desses animais, utilizando o enriquecimento ambiental. De acordo com Martins (2017) no alojamento de camundongos em biotérios é primordial o conhecimento do comportamento desses animais e o uso de enriquecimento ambiental para elevar a condição de bem-estar e minimizar o estresse durante o período de manutenção e uso dos animais.

Além do desperdício dos animais produzidos em excesso, o descarte apresenta prejuízos financeiros. Durante a pesquisa interna no biotério, citada anteriormente, através do relatório foi realizada a precificação do custo para produzir um único camundongo pelo LACEA, sendo ele no valor de R\$106,17 (reais). No ano de 2019 até o mês de junho, foram R\$ 124.218,90 (reais) em camundongos descartados, causando desperdício de verba pública.

O presente trabalho, em que foram descritos os parâmetros de produção animal e descarte, além de encorajar mudanças de gerenciamento da colônia, provocou uma discussão sobre os princípios dos 3 Rs (*The Principles of Humane Experimental Technique - Replace, Reduce e Refine*), principalmente no quesito redução. Após os resultados obtidos em nossas análises, a equipe do LACEA demonstra novos planejamentos, como a redução do número de casais, evitando manter animais de prateleira (que são produzidos e ficam aguardando solicitação), buscas ativas de clientes para comercialização de lotes de camundongos fêmeas /ou macho que não são aproveitados da produção dos pedidos internos. Camundongos produzidos em excesso são pré desmamados no período neonato e descartados com

poucos dias de vida, reduzindo a manutenção de animais. Além de outras medidas para reduzir o descarte de animais.

Para a nova versão do BioterC (BioterCS), foi sugerido incluir recursos de Aprendizado de Máquina (do inglês, *Machine Learning*) que tem como objetivo uma análise avançada, permitindo que as empresas abandonem a análise descritiva e voltem-se para o suporte à decisão autônoma. Praticamente qualquer setor da indústria pode utilizar o aprendizado de máquina e da análise avançada. Sempre que grandes volumes de dados e modelos preditivos necessitam de ajustes regulares, essa técnica permite realizar de modo automatizado (não dependem de intervenção ou interação humanas para solucionar uma nova classe de decisões) (Intel, 2019). Essa ferramenta pode ser aplicada para previsão de fornecimento de animais, previsão de montagem de acasalamentos, entre outras etapas da criação animal gerenciadas por esse *software*.

6. Conclusões

Com base em nossos achados, foi possível comparar os padrões de criação de camundongos com o que já é observado na prática, bem como encontrado na literatura, corroborando com nossa hipótese. Além disso, a partir dos resultados obtidos fomentaram mudanças nas rotinas do LACEA, demonstrando melhorias no gerenciamento do plantel através de novos planejamentos internos do biotério.

Também, foi possível observar o excesso de descarte de animais que não são aproveitados durante o experimento, abrindo a discussão na comunidade acadêmica sobre o excesso do uso de animais em pesquisas científicas, a partir dos princípios dos 3 R's, contribuindo para Substituir, Reduzir e Refinar.

A utilização da *Data Science* para trabalhar com conjunto de dados do BioterC, garantiu vantagens na exploração das informações de gerenciamento da criação animal, com agilidade e otimização no processamento e estruturação de informações. Outra vantagem, foi de não precisar recorrer a soluções pagas, apenas o conhecimento prévio de linguagem de programação. Além de gerar resultados confiáveis através de cálculos matemáticos e estatísticos.

7. Referências

BENCHIMOL, J. L. **Manguinhos do sonho à vida - A ciência na Belle Époque**. Ed: Casa de Oswaldo Cruz. Rio de Janeiro, 1990.

BENAVIDES, F.J.; GUENÉT, J.L. **Manual de Genética de Roedores de Laboratório: Princípios básicos y aplicaciones**. Madri: Universidad de Acalá, SECAL, 2003. p 63-65.

BOX, G.E.P., JENKINS, G.M. AND REINSEL, G.C. **Time Series Analysis; Forecasting and Control**. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey, 1994.

BRAGA, L. M. Controle reprodutivo em biotérios de criação de animais de laboratório com ênfase em roedores - **Rev. Bras. Reprod. Anim.**, Belo Horizonte, 2017 v.41, n.1, p.105-109.

BROWN, R., MATHIESON B., STAPLETON J., NEUMANN P. Maternal Behavior in Female C57BL/6J and DBA/2J Inbred Mice - **Physiology & Behavior** – vol 67. p 599-605. Issue 4, outubro, 1999.

CARDOSO, C. V. **Criação e manejo de camundongos** – Cap 30 Eutanásia. In: Andrade A, Pinto SC, Oliveira RS. Animais de laboratório: criação e experimentação. Rio de Janeiro: Fiocruz, 2002. b p.115-8.

CARDOSO T.A.O. Considerações sobre a biossegurança em arquitetura de biotérios. **Bol. Centr. Panam. Fiebre Aftosa**, 2001; 64-67:3-17

CASELLAS J. Inbred mouse strains and genetic stability: a review. *Animal*. 5:1, 2010 p1-7.

CCAC. History about Canadian Council on Animal Care. CCAC Disponível em < <https://ccac.ca/en/about-the-ccac/>>. Acesso em 09 mar de 2021.

CDC. Centers for Disease Control. Laboratory safety. Atlanta, US: Dep. Health, Education and Welfare/**Public Health Service**; 1979. (HEW Publication, CDC79-818).

CHATFIELD, C. **Time Series Forecasting**, Chapman & Hall/CRC, 2000.

CONCEA. **Euthanasia Practice Guidelines** - CEUA/IOC 2013.

CRUZ, C. A. **Universitas Ciências da Saúde** - Brasília - vol.01 n.02 - pp.343-362, 2003.

DALBEN, D.; EMMEL, J. L. A lei Arouca e os direitos dos animais utilizados em experimentos científicos. **Revista Eletrônica de Iniciação Científica**. Itajaí, Centro de Ciências Sociais e Jurídicas da UNIVALI. v. 4, n.4, p. 280-291, 2013.

DIRETRIZES BIOTÉRIO. Diretrizes do laboratório de criação e experimentação animal (LACEA) – FIOCRUZ/PR, 2017.

DUARTE GI. **Programa de Ação para Biotérios** – Rede Nordeste de Biotérios, Centro de Gestão e Estudos Estratégicos (CGEE). São Paulo; 2003.

FEKETE, E. In G.D. Snell [ed.] **Biology of the Laboratory Mouse**. Blakiston, Philadelphia., 1941. p 56

FELASA. Federation of European Laboratory Animal Science Associations. FELASA guidance paper for the accreditation of laboratory animal diagnostic laboratories. **Laboratory Animals**. 1999; 33(suppl 1):19-51.

FELASA. Federation of European Laboratory Animal Science Associations. FELASA recommendations for the health monitoring of mouse, rat, hamster, guinea pig and

rabbit colonies in breeding and experimental units. **Laboratory Animals** 2014, Vol. 48(3) 178–192

FESTING MF. **Genetic variation in outbred rats and mice and its implications for toxicological screening.** J Exp Anim Sci. 1993; 35(5-6):210-20.

FIOCRUZ. Fundação Oswaldo Cruz. Comissão Técnica de Biossegurança (CTBio). **Procedimentos para a manipulação de microorganismos patogênicos e/ou recombinantes na FIOCRUZ.** Rio de Janeiro: Fundação Oswaldo Cruz; 1998.

GUÉNET J.L., BENAVIDES F., PANTHIER J.J., MONTAGUTELLI X. Genetics of the Mouse. Berlin: **Springer**, 2015 p.408.

JONAS AM. The research animal and the significance of a health monitoring program. **Lab Anim Sci.** 1976; 26(2):339-43.

KEELER, C.E. **The Laboratory Mouse: Its Origin, Heredity, and Culture.** Harvard University Press, Cambridge, Mass, 1931.81 p.

KIKUSI, T. *et al.* Early weaning deprives mouse pups of maternal care and decreases their maternal behavior in adulthood. **Behavioural Brain Research.** Volume 162;162:200-206. Issue 2, 30 July, 2005.

KO, G. M. IN: LAPCHIK, V. B. V.; MATTARAIA, V., G., M.; KO, G., M. (Ed.). **Cuidados e manejo de animais de laboratório.** São Paulo: Atheneu, 2009. p. 137-164.

KRACKER, L. FERRO, P. JR CONSULTORIA. Relatório Interno Biotério para– Instituto Carlos Chagas. UFPR, 2019 p. 7

INTEL. Machine learning: o próximo passo em análise avançada. 2019. Disponível em < <https://www.intel.com.br/content/www/br/pt/company-overview/contact-us.html>> Acesso em 11 mar 2021.

LAPCHIK, V. B. V.; MATTARAIA, V. G. M.; KO, G. M. **Cuidados e Manejo de Animais de Laboratório**. São Paulo:EditoraAtheneu, 2010.

LINDEN A., ADAMS J.L., ROBERTS N. Evaluating disease management program effectiveness: an introduction to time series analysis. *Dis Manag.* 2003;6(4):243–255.

MAJEROWICZ J. **Procedimentos de biossegurança para as novas instalações do laboratório de experimentação animal (Laean) de Bio-Manguinhos**. [Dissertação] Rio de Janeiro: Instituto Oswaldo Cruz, Instituto de Tecnologia em Imunobiológicos; 2005.

MARTINS, Thais Veronez de Andrade. **Avaliação da preferência pelo tipo de enriquecimento ambiental utilizado por camundongos swiss webster em biotério através do sistema de gaiolas interligadas (SGI)**. 2017. 67 f. Dissertação (Mestrado em Ciência em Animais de Laboratório) -Instituto de Ciência e Tecnologia em Biomodelos, Fundação Oswaldo Cruz, Rio de Janeiro, 2017.

MASSIRONI, S. M. G. Padrão genético. In: LAPCHIK, V. B. V; MATTARAIA, V. G. M; KO, G. M. **Cuidados e Manejo de Animais de Laboratório**. Atheneu, 2009 p. 385 – 398.

MAZZAROTTO, G. A. C. A., SILVEIRA, G. F. Desenvolvimento e implementação de um software livre para o gerenciamento de um biotério brasileiro - **RESBCAL**, São Paulo v.2 n.1, 2013. p 61-68.

MORETTIN, PEDRO A. E TOLOI, CLÉLIA M. Modelos para Previsão de Séries Temporais, Edgard Blucher, 1981.

NCR3RS. National Centre for Replacement Refinement & Reduction of Animals in Research. The 3Rs. Disponível em: <<https://www.nc3rs.org.uk/the-3rs>> Acesso em 15 abr 2021.

NEVES, S. M. P, MANCINI, J. F, MENEZES, E.W. **Manual de Cuidados e Procedimentos com Animais de Laboratório do Biotério de Produção e Experimentação** da FCF-IQ/USP. 2013. P.11.

NIH. National Institute of Health. **Environment, housing and animal management. In: Institutional Animal Care and use Committee.** Book guidebook. 2nd. ed. Bethesda: NIH; 2002. (Publication n°. 92-3415).

NIH. National Institute of Health. **Environment, housing and animal management. In: Institutional Animal Care and use Committee** Book guidebook. 2nd. ed. Bethesda: NIH; 2002. (Publication n°. 92-3415).

NRC.National Research Council (US) and Institute of Medicine (US) Committee on the Use of Laboratory Animals in Biomedical and Behavioral Research. **Use of Laboratory Animals in Biomedical and Behavioral Research.** Washington (DC): National Academies Press (US); 1988. 2, Patterns of Animal Use. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK218261/>

NRC. National Research Council. Institute of Laboratory Animal Resources. Commission on life sciences. Manual sobre Cuidados e Usos de Animais de Laboratório. Goiânia: **AAALAC/COBEA**, 2003. 162p.

PAIVA F., MAFFILI V., SANTOS A. Curso de manipulação de animais de laboratório. Salvador: Fundação Oswaldo Cruz (FIOCRUZ), Centro de Pesquisas Gonçalo Moniz, 2005.

PASSOS LAC. **Análise do determinismo genético da resistência de camundongos infectados experimentalmente com a cepa y de *Trypanosoma cruzi*.** Campinas; 2003. [Tese] Campinas: Instituto de Biologia. Universidade Estadual de Campinas; 2003.

PRINS, J. Federation of European Laboratory Animal Science Associations (FELASA) 2018.

PROVOST F., FAWCETT T., 2013. Ciência de dados and its relationship to big data and data-driven decision making. <<https://www.liebertpub.com/doi/pdfplus/10.1089/big.2013.1508>> Acesso em: 05 ago 2019.

SALGADO, H.R.N; POLITI, F.A.S.; MAJEROWICZ, J.; CARDOSO, T.A.O.; PIETRO, R.C.L.R.; Caracterização de biotérios, legislação e padrões de biossegurança. **Rev. Ciênc. Farm. Básica Apl.**, vol. 29, n.1, p. 17-28, 2008 ISSN 1808-4532

SANTOS, BF. **Criação e manejo de camundongos**. In: Andrade A, Pinto SC, Oliveira RS. Animais de laboratório: criação e experimentação. Rio de Janeiro: Fiocruz, 2002. p.115-8.

SIMMONS, J.G. **Médicos e descobridores – vidas que criaram a Medicina de hoje**. São Paulo: Record, 2004.

WIENER N. **Extrapolation, interpolation and smoothing of stationary time series**. Cambridge: MIT Press; 1966.

WYLINKA. Software para Controle Gerencial de Biotérios - BIOTERC. Projeto de Avaliação e Comercialização de Novas Tecnologias. Instituto Carlos Chagas, 2016.

8. Apêndice

8.1. Quadros apresentando funções e termos computacionais

Quadro 8.1 – Termologias utilizadas em linguagens de programação

Python	Linguagem de programação interpretada e orientada a objetos.
Anaconda	Ambiente de hospedagem para criação de códigos de distribuição gratuita e de código aberto para computação científica (ciência de dados), que visa simplificar o gerenciamento de pacotes e desdobramento, desenvolvimento.
IDE	programa que reúne características e ferramentas de apoio ao desenvolvimento de <i>software</i> com o objetivo de apoiar o processo.
SQL	sistema gerenciador de banco de dados relacional, que tem como principal função permitir o armazenamento e a recuperação estruturada de informação
DataFrame	É um quadro de dados com uma estrutura bidimensional, ou seja, os dados são alinhados de forma tabular em linhas e colunas. O Pandas <i>DataFrame</i> consiste em três componentes principais, os dados, linhas e colunas
Jupyter Lab	Projeto Jupyter existe para desenvolver softwares de código aberto e serviços para computação interativa em dezenas de linguagens de programação.
Bitbucket/Github	É um serviço de hospedagem de projetos

8.2. Métodos utilizados para Análise Exploratória de Dados

Quadro 8.2- Biblioteca Pandas de Python

<code>read_csv()</code>	Leia um arquivo de valores separados por vírgula (csv) no <i>DataFrame</i> .
<code>groupby()</code>	Uma operação de agrupamento envolve alguma combinação de divisão do objeto, aplicação de uma função e combinação dos resultados.
<code>resample()</code>	Método de conveniência para conversão de frequência e reamostragem de séries temporais. O objeto deve ter um índice semelhante ao <i>datetime</i> .
<code>isin()</code>	Se cada elemento no <i>DataFrame</i> está contido em valores.
<code>lambda</code>	Permite criar outra função que será utilizada apenas uma vez
<code>sum()</code>	Calculando soma.
<code>median()</code>	Calculando mediana.
<code>Values.count()</code>	Organiza e retorna o número de instâncias do maior para o menor.

Quadro 8.3 - Biblioteca Numpy em Python

<code>quantile()</code>	Retorne valores no quartil fornecido sobre o eixo solicitado.
-------------------------	---

Quadro 8.4 - Biblioteca *Statsmodels* em Python

<code>seasonal_decompose()</code>	Classe tsa: Decomposição sazonal usando médias móveis.
-----------------------------------	--

Quadro 8.5 - Biblioteca *Seaborn* e *Matplotlib* em Python

<code>sns.distplot()</code>	Plotando gráfico do tipo histograma.
<code>boxplot()</code>	Plotando gráfico do tipo boxplot.
<code>plot()</code>	Plotando diferentes gráficos.

9. Artigo

Ferreira et al. *Laboratory Animal Research* (2021) 37:1
<https://doi.org/10.1186/s42826-020-00082-w>

Laboratory Animal Research

RESEARCH

Open Access

Zootechnical data analysis in a breeding animal facility: tracing the patterns of mouse production



Eloiza K. G. D. Ferreira¹, Giovanni A. C. A. Mazzarotto^{1,2} and Guilherme F. Silveira^{1*}

Abstract

Background: With the enactment of the Brazilian Law Arouca 11,794/2008 and Decree 6.899/2009, there has been an urgent need for changes in the processes related to animal experimentation in Brazil; in particular, there is a need for improvements in enhancements of the lab animal management. To improve the management capacity of the Lab animal facility of the Carlos Chagas Institute's Laboratory Animals Science (LAS), BioterC software was developed and implemented in 2014 for tracking mouse laboratory colonies. Five years after the implementation of this software, we sought to analyze the information in the database originated from BioterC using the Exploratory Analysis Data methodology (EDA). This article aims to identify animal breeding patterns using a data mining tool (Data Science) with Python programming language.

Results: The results show that from September 2014 to June 2019, under the license IACUC number LW- 6/17, 15,106 animals were produced. The C57BL/6, BALB/c and Swiss strains were the most frequently produced strains. The distribution of births due to crosses between these strains showed a median of 6 to 10 animals, depending on the genetic homozygosis and heterozygosis of the animal. The median number of days of mating was 35 days. In the sexing period, the records reported a median of 19 days. A total of 393 requests for animals from internal and external laboratories were registered. It was noted that approximately half of the animals produced to meet the demand for orders were discarded. Of the 15,106 animals, 38% were requested for animal experimentation, 58% were discarded and 4% did not have an outcome recorded in the data.

Conclusions: This volume of data provides an initial view of the information retrieval capabilities contained in BioterC, allowing for unique breeding knowledge by installing laboratory animals.

Keywords: Data science, Exploratory data analysis (EDA), Python, Laboratory animals science (LAS)

Background

History of the animal model

The use of animals as a living model to achieve scientific aims has been common practice since the seventeenth century. Initially, these models were used for studies of anatomy and physiology involving dissection to analyze body functions [1].

The first Brazilian laboratory animal facility that produced animals for research was inaugurated in 1904 at the Sorotherapeutic Institute of Rio de Janeiro, which today is the current Oswaldo Cruz Foundation [2]. The changes in scientific research conduct and technological advances observed in the last 116 years have led to the need for improvements in facilities due to the increasing use of animals for teaching and research.

* Correspondence: gfsilveira@gmail.com

¹Instituto Carlos Chagas- Fiocruz/PR, Curitiba, PR, Brazil

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.