
CIÊNCIAS BIOLÓGICAS

JESSICA SAYURI SAITO

**ESTUDO DE TÉCNICAS DE ESTATÍSTICA,
VISANDO APLICAÇÕES A DADOS DE
NATUREZA BIOLÓGICA**



Rio Claro
2020

JESSICA SAYURI SAITO

ESTUDO DE TÉCNICAS DE ESTATÍSTICA, VISANDO APLICAÇÕES A
DADOS DE NATUREZA BIOLÓGICA

Orientador: José Silvio Govone

Trabalho de Conclusão de Curso apresentado ao Instituto de Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho” - Câmpus de Rio Claro, para obtenção do grau de Licenciada e Bacharela em Ciências Biológicas.

Rio Claro
2020

S158e Saito, Jessica Sayuri
Estudo de técnicas de estatística, visando aplicações a dados de natureza biológica / Jessica Sayuri Saito. -- Rio Claro, 2020
98 f. : tabs.

Trabalho de conclusão de curso (Bacharelado e licenciatura - Ciências Biológicas) - Universidade Estadual Paulista (Unesp), Instituto de Biociências, Rio Claro
Orientador: José Silvio Govone

1. Bioestatística. 2. População Estatísticas. 3. Mosca-das-frutas. 4. Mosca-varejeira. 5. Análise de sobrevivência (Biometria). I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

*Dedico este trabalho
aos meus pais*

AGRADECIMENTOS

Agradeço a minha família por apoiar minhas decisões, incentivar a continuar o caminho escolhido, e ajudar nos momentos mais difíceis, mesmo que estivesse longe de casa.

Agradeço a minha querida amiga Ana Vitória e minha gatinha Nina por sempre estarem comigo durante estes 5 anos, fornecendo apoio emocional nos momentos mais felizes e mais difíceis da graduação. Agradeço também as minhas amigas de curso Giselle, Isabela, Juliana e Tamiris por sempre me ajudarem com os trabalhos, as tarefas e a lidar com problemas de estresse decorrente da graduação. Sempre guardarei memórias de minhas amigas com muito carinho, e espero manter contato com elas nos anos que seguirão.

Agradeço ao Prof. Dr. José Silvio Govone pela oportunidade e pelo todo o tempo que dispôs, com muita dedicação, para me supervisionar e orientar durante o estágio obrigatório e o TCC. Sempre contarei com o senhor quando o assunto é estatística. Agradeço também ao Prof. Dr. Cláudio José Von Zuben pelo fornecimento dos dados que foram analisados neste estudo, assim como pelo apoio ao meu trabalho.

Apóie-se nos seus medos, incite-os a fazer o pior possível e derrube-os quando tentarem. Se não o fizer, eles se multiplicam, envolvem-no e sufocam pela a estrada da vida que você quer seguir.

Toda a curva que você teme é ar vazio, disfarçado para aparentar um demônio (BACH, 2006, p. 28).

RESUMO

Neste trabalho, inicialmente realizou-se um estudo teórico sobre análise exploratória de dados, testes de normalidade e homocedasticidade, testes paramétricos e não paramétricos para comparações de grupos, transformações de dados, análise de sobrevivência e análise de componentes principais. Buscando aplicar o conhecimento adquirido, foram realizadas análises nos conjuntos de dados referentes ao número de coletas de moscas-das-frutas, dados sobre postura de califorídeos da espécie *Chrysomya megacephala*, fornecidos pelo Prof. Dr. Claudio J. Von Zuben, Departamento de Biodiversidade (ex-Zoologia), assim como em dados necrológicos de 2019 do município de Bebedouro, SP. Nas análises do primeiro conjunto de dados, os drosofilídeos foram agrupados por gêneros e expostos aos testes de comparações de grupos, análise de proporção e correlação, esperando-se verificar diferenças significativas na abundância de moscas no decorrer dos meses e estações, assim como a presença de associações entre os gêneros. Na análise do segundo conjunto de dados, foi aplicado teste de comparações de grupos para averiguar a ocorrência de postura agregada nos potes, os quais continham o mesmo substrato em posições distintas, e estavam contidos em gaiolas (réplicas) com diferentes densidades de califorídeos. No terceiro conjunto de dados realizou-se a análise gráfica das curvas de sobrevivência, frequências, probabilidades e taxas de falhas nas populações feminina e masculina de Bebedouro, assim como teste de hipóteses, buscando verificar diferenças em suas sobrevivências. Os resultados da primeira análise revelaram diferenças significativas entre meses, decorrentes do maior número de coletas de drosofilídeos em períodos quentes e úmidos, que comumente estimulam a frutificação da figueira, assim como o desenvolvimento e reprodução de moscas. Nesta primeira parte também se verificou uma possível coexistência estável entre os gêneros de drosofilídeos, ao invés de competição interespecífica, pela existência de correlações significativas positivas. Por outro lado, observaram-se diferenças entre este trabalho e a literatura quanto à flutuação da abundância destes insetos, as quais podem refletir as particularidades das condições ambientais e da comunidade de seres vivos dos locais estudados, que influenciam na composição e nos nichos ecológicos ocupados pelos drosofilídeos. No segundo conjunto, os resultados revelaram apenas diferenças significativas nas quantidades de posturas de *C. megacephala* em 45,45% das réplicas, as quais se concentraram naquelas de menores densidades, corroborando com a teoria de avaliação do substrato pela fêmea do califorídeo. As análises do terceiro conjunto de dados evidenciaram probabilidades e taxas de falhas elevadas para as idades mais avançadas, assim como, diferenças nas curvas de sobrevivência entre os sexos, averiguando-se uma

sobremortalidade masculina em idades anteriores a 60 anos, a qual a literatura sugere que seria devido a causas não naturais (acidentes, violência, entre outros fatores).

Palavras-chave: Bioestatística. Testes de comparação de grupos. Moscas-das-frutas. *Chrysomya megacephala*. Análise de sobrevivência.

SUMÁRIO

1 INTRODUÇÃO	9
1.1 Análises comuns em bioestatística	9
<i>1.1.1 Análise exploratória de dados</i>	10
<i>1.1.2 Testes de normalidade</i>	12
1.1.2.1 <u>Análise gráfica</u>	12
1.1.2.2 <u>Teste Shapiro-Wilk</u>	12
<i>1.1.3 Testes de homogeneidade de variâncias</i>	13
1.1.3.1 <u>Análise gráfica</u>	13
1.1.3.2 <u>Teste de Levene</u>	13
<i>1.1.4 Testes de hipóteses para comparação de grupos</i>	14
1.1.4.1 <u>Análise de variância (ANOVA)</u>	15
<i>1.1.4.1.1 <u>ANOVA de blocos aleatorizados</u></i>	16
1.1.4.2 <u>Teste de Kruskal-Wallis</u>	17
1.1.4.3 <u>Teste de Friedman</u>	17
<i>1.1.5 Correlação</i>	18
1.1.5.1 <u>Correlação de Spearman</u>	18
<i>1.1.6 Análise de sobrevivência</i>	19
1.1.6.1 <u>Estudos censitários</u>	19
1.1.6.2 <u>Estudos amostrais</u>	19
<i>1.1.6.2.1 <u>Kaplan-Meier</u></i>	20
<i>1.1.6.2.2 <u>Log-Rank test</u></i>	21
1.2 Dípteros como modelo biológico	21
<i>1.2.1 Moscas-das-frutas</i>	22
<i>1.2.2 <i>Chrysomya megacephala</i></i>	23
1.3 Objetivos	24
2 MATERIAIS E MÉTODOS	25
3 RESULTADOS E DISCUSSÃO	27
3.1 Análise do conjunto de dados de moscas-das-frutas agrupadas em gêneros	27
<i>3.1.1 Comparação do número de coletas de moscas-das-frutas entre os meses</i>	27
<i>3.1.2 Comparação do número de coletas de moscas-das-frutas entre estações</i>	32
<i>3.1.3 Correlação entre os gêneros de moscas-das-frutas</i>	37

3.1.4 Análise gráfica de frequências das coletas de moscas-das-frutas.....	40
3.2 Análise dos dados referentes às posturas de <i>C. megacephala</i> em diferentes densidades.....	45
3.2.1 Posturas de <i>C. megacephala</i> na densidade 100 califorídeos por réplica.....	45
3.2.2 Posturas de <i>C. megacephala</i> na densidade 200 califorídeos por réplica.....	48
3.2.3 Posturas de <i>C. megacephala</i> na densidade 400 califorídeos por réplica.....	50
3.2.4 Posturas de <i>C. megacephala</i> na densidade 800 califorídeos por réplica.....	54
3.3 Análise de sobrevivência dos dados de Bebedouro 2019.....	59
4 CONCLUSÃO.....	66
5 REFERÊNCIAS.....	67
APÊNDICE A - TESTE DE BARTLETT.....	75
APÊNDICE B - TESTE T-STUDENT.....	76
APÊNDICE C – ANOVA.....	78
APÊNDICE D- TRANSFORMAÇÃO DE DADOS.....	82
APÊNDICE E- CORRELAÇÃO.....	83
APÊNDICE F- REGRESSÃO SIMPLES E MÚLTIPLA.....	86
APÊNDICE G- ANÁLISE DE SOBREVIVÊNCIA.....	89
APÊNDICE H- ANÁLISE DE COMPONENTES PRINCIPAIS.....	91
APÊNDICE I- RESULTADOS DA ANÁLISE DE SOBREVIVÊNCIA DO MUNICÍPIO BEBEDOURO, SP.....	92

1 INTRODUÇÃO

A estatística constitui-se de uma ciência exata que apresenta conjuntos de métodos, os quais auxiliam no planejamento, obtenção, organização, resumo e análise de dados, possibilitando a obtenção de resultados e conclusões com maior grau de confiabilidade em estudos das diferentes áreas do conhecimento, como na saúde, economia e ciências da natureza. Além de disponibilizar uma sequência de métodos, a estatística estimula a obtenção de postura crítica pelo pesquisador no momento da análise e interpretação dos resultados, possibilitando a descrição, explicação ou determinação de prováveis implicações do fenômeno estudado (AYRES, 2011; CALLEGARI-JACQUES, 2003).

Segundo Ayres (2011), a estatística pode ser dividida em estatística matemática e aplicada. A primeira consiste dos conjuntos de métodos matemáticos, produzidos ao longo do tempo por diversos estudiosos, enquanto a estatística aplicada utiliza destes métodos para elucidar problemas de diversas áreas do conhecimento (AYRES, 2011).

A estatística aplicada pode ser descritiva ou inferencial, sendo que a primeira realiza o levantamento de dados de uma amostra ou de uma população, organizando e descrevendo-os, através de medidas de posição, dispersão ou por meio de tabelas e gráficos, de maneira a possibilitar a análise e a obtenção de conclusões. Diferentemente, a inferência estatística constitui-se da generalização de um fenômeno específico para uma população, podendo ser acompanhada de exceções e de margem de incerteza (AYRES, 2011; CALLEGARI-JACQUES, 2003).

Quando a estatística é utilizada para resolução de problemas biológicos, esta passa a ser chamada de bioestatística. A bioestatística possibilita uma interpretação adequada e compreensão da diversidade e dinamicidade de fenômenos biológicos por meio de testes probabilísticos, de hipóteses e modelos computacionais, ao considerar que existe variabilidade intraespecífica dentro de populações de seres vivos, sendo que seus indivíduos podem também sofrer variações de um momento para o outro (CALLEGARI-JACQUES, 2003; GOTELLI; ELLISON, 2011).

1.1 Análises comuns em bioestatística

Nesta primeira seção apresenta-se um breve resumo dos métodos estatísticos utilizados neste trabalho. Os resumos dos outros métodos estudados apresentam-se disponíveis nos apêndices.

1.1.1 *Análise exploratória de dados*

Após a obtenção dos dados através de fontes primárias (coletada pelo próprio pesquisador) ou secundárias (adquiridas através de bancos de dados), deve-se realizar a análise exploratória destes dados, ou como chamada antigamente, a estatística descritiva. Os dados coletados podem ser referentes a censos, ou à amostragem representativa de uma determinada população. Quando se trata de um estudo amostral, esta pode se constituir de uma pesquisa observacional ou experimental, na qual a última permite o controle das covariáveis e variáveis de interesse (AYRES, 2011; GOTELLI; ELLISON, 2011).

A análise exploratória de dados tem como objetivo a organização e resumo de dados, por meio de tabelas (frequências absolutas ou relativas, porcentagem ou contingência), gráficos (dispersão, barras, setores, *box-plot* ou histograma), da estimação de parâmetros populacionais acerca da posição de seus valores (média, separatrizes e moda), dispersão (variância, desvio padrão, erro padrão e coeficiente de variação) e de suas tendências de distribuição (assimetria e curtose), buscando facilitar a identificação de padrões e interpretação das observações. As separatrizes ou quantis constituem-se de medidas que dividem os dados em partes iguais, como mediana (2 partes), quartis (4 partes), decis (10 partes) e centis (100 partes). A assimetria (g_1) e a curtose (g_2) referem-se respectivamente a parâmetros de distribuição dos dados e das suas medidas de posição, e sobre a densidade de probabilidades das observações das caudas em relação ao seu centro (GOTELLI; ELLISON, 2011; SOKAL; ROHLF, 1995).

Quando o conjunto de dados apresenta distribuição simétrica ($g_1=0$), os parâmetros de posição, média, mediana e moda possuem valores idênticos. Quando ocorre assimetria, esta pode ser positiva ou negativa. A assimetria positiva ou à direita ($g_1>0$), ocorre quando a média apresenta maior valor comparado à mediana, devido a grande parte dos dados concentrarem no início do gráfico, o qual exibe uma longa cauda com poucas observações mais elevadas. Na assimetria negativa ou à esquerda ($g_1<0$) verifica-se uma média inferior à mediana, em que os dados estão concentrados no final do gráfico, o qual mostra uma longa cauda com poucas medidas de valores reduzidos (AYRES, 2011; GOTELLI; ELLISON, 2011).

Quando a distribuição das probabilidades dos valores se aproxima de uma curva normal gaussiana, esta apresenta uma curtose do tipo mesocúrtica ($g_2=3$). Quando as massas de probabilidades são mais densas nas caudas em relação ao centro da curva, ocorre uma

curtose do tipo leptocúrtica ($g_2 < 3$). Por outro lado, a curtose do tipo platicúrtica ($g_2 > 3$) é caracterizada por apresentar maior concentração das massas de probabilidades dos dados na região central do gráfico em relação às caudas, quando comparado com a curva leptocúrtica (GOTELLI; ELLISON, 2011).

Esta primeira etapa da análise de dados também permite verificar a existência de dados faltantes (*missing values*), dados discrepantes (medidas inferiores ao 5º ou superiores ao 95º percentil, não se encontrando nos intervalos da maioria das observações) e erros, por meio da avaliação das informações tabeladas e pela análise gráfica (principalmente por gráficos de dispersão e *box-plot*). A identificação destes dados possibilita a posterior correção desses valores pela: substituição pela média do grupo ou por outra observação; incorporação desses dados como parte de sua variabilidade; ou simplesmente o descarte, principalmente quando o valor corresponde a um erro (não representa as relações entre as variáveis). O descarte deve ser evitado para os outros casos, considerando-se a perda de informações decorrente desta ação, que pode prejudicar o estudo do fenômeno de interesse (AYRES, 2011; GOTELLI; ELLISON, 2011).

Além disso, realiza-se também nesta primeira etapa a identificação e a classificação das variáveis de interesse segundo o objetivo do estudo. As variáveis podem ser classificadas em qualitativas ou quantitativas, e independentes ou dependentes. As variáveis qualitativas ou categóricas são definidas por grupos ou categorias, podendo ser subdivididas em nominais (não ordenadas, como o gênero do indivíduo) e ordinais (ordenadas, como as classes socioeconômicas). Por outro lado, as variáveis quantitativas são numéricas, podendo ser discretas (assumem valores determinados, como o número de indivíduos) ou contínuas (assumem amplitude de valores, como a massa de um objeto). Quanto ao nível de manipulação, as variáveis podem ser classificadas em independentes (preditoras) ou dependentes (respostas), sendo as últimas influenciadas pela alteração dos valores da variável preditora (GOTELLI; ELLISON, 2011).

A classificação inicial das variáveis possibilita direcionar a análise mais adequada. A análise de regressão é recomendada quando as variáveis preditora e resposta são numéricas contínuas, exceto na regressão logística, na qual a variável dependente é categórica, assumindo valores binários. Por outro lado, quando a variável dependente é categórica e a independente numérica contínua, prefere-se utilizar o delineamento de ANOVA (GOTELLI; ELLISON, 2011).

1.1.2 Testes de normalidade

A distribuição normal dos dados e dos seus resíduos constitui-se de pressuposto comum em testes de hipóteses paramétricos e de análises de regressão, sendo indispensável a verificação desta condição por meio da aplicação de análises gráficas e testes de hipóteses para a verificação da hipótese nula (H_0) acerca da normalidade das observações (AYRES, 2011; ZAR, 1999).

1.1.2.1 Análise gráfica

A construção de gráficos do tipo *QQplots* permite verificar a normalidade dos dados por meio da representação gráfica de uma reta, quando os quantis dos dados são dispostos contra seus quantis teóricos ($q(f_i)$) (PORTAL ACTION, c2020a; UFMG, 2011).

Os quantis teóricos podem ser estimados a partir de f_i , que se constitui de uma estimativa da proporção de valores menores que cada um dos dados originais ordenados em ordem crescente $x_{(i)}$, de uma amostra $i=1,2,...,n$ observações (UFMG, 2011).

$$f_i = \frac{i - a}{n + 1 - 2a} \quad (1.01)$$

Sendo que a representa o fator de correção da equação. Quando n é menor ou igual a 10, $a=3/8$, por outro lado, quando $n > 10$, a assume valor de $1/2$ (PORTAL ACTION, c2020a; UFMG, 2011).

1.1.2.2 Teste Shapiro-Wilk

O teste de Shapiro-Wilk consiste de um teste de normalidade, o qual se baseia na estatística W , cujo valor pode ser calculado através da formula a seguir:

$$W = \frac{b^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.02)$$

Sendo que x_i representa os valores das $i=1,2,...,n$ observações organizadas em ordem crescente; $\bar{x}$ a média das observações; e b uma constante calculada pela equação abaixo (PORTAL ACTION, c2020a; UEL, 201-?a).

$$b = \sum_{i=1}^k a_{n,i} (x_{n-i+1} - x_i) \quad (1.03)$$

Em que $a_{n,i}$ consiste em uma constante tabelada para tamanho amostral n (obtida a partir das médias, variâncias e covariâncias de n); k corresponde a metade dos valores amostrais, sendo $k=n/2$, para n par, e $k=(n+1)/2$, para n ímpar; e $x_{(n-i)+1}$ representa a medida complementar da observação x_i (PORTAL ACTION, c2020a; UEL, 201-?a).

Por meio do valor de W calculado, compara-se com o valor W_{tab} , (linha n , coluna α da tabela W), aceitando a hipótese nula sobre a distribuição normal dos dados quando $W \geq W_{tab}$, ou, equivalentemente, o valor de probabilidade p do teste for maior ou igual a probabilidade de erro α fixada. (PORTAL ACTION, c2020a; UEL, 201-?a).

Além da análise gráfica e o teste Shapiro-Wilk, outros testes podem ser utilizados para testar a normalidade dos dados, como o teste de Kolmogorov-Smirnov, teste de Anderson-Darling e teste de Ryan-Joiner (PORTAL ACTION, c2020a).

1.1.3 Testes de homogeneidade de variâncias

Além da distribuição normal, alguns testes de hipóteses paramétricos também apresentam como pressuposto a igualdade das variâncias dos dados ou, equivalentemente, de seus resíduos (homocedasticidade). Esta condição pode ser verificada através de análise gráfica e de testes que avaliem a hipótese nula acerca da distribuição igual das variâncias ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$) (AYRES, 2011; ZAR, 1999).

1.1.3.1 Análise gráfica

Comumente utiliza-se o gráfico de resíduos em relação aos valores preditos ou ajustados para o diagnóstico de homocedasticidade, verificando o cumprimento deste pressuposto pela distribuição aleatória de seus pontos ao redor do marco zero. Por outro lado, quando se observa uma tendência na distribuição dos pontos, pode ser indicativo de heterocedasticidade e associação entre as observações (PORTAL ACTION, c2020b).

1.1.3.2 Teste de Levene

O teste de Levene é comumente utilizado para avaliar a igualdade de variâncias de conjunto de dados que não apresentam distribuição normal. Esse teste constitui-se da aplicação do teste ANOVA aos dados transformados (baseado na estatística F), como apresentado na equação a seguir:

$$F^* = \frac{\sum_{i=1}^k \frac{n_i(\bar{Z}_i - \bar{Z})^2}{(k-1)}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2}{\sum_{i=1}^k (n_i - 1)}} \quad (1.04)$$

Em que Z_{ij} , $\bar{Z}_i$ e $\bar{Z}$ correspondem a cada observação transformada de determinado nível do fator $i=1,2,...,k$ e réplica $j=1,2,...,n$, a média dos dados transformados de um nível i , e a média geral, respectivamente (PORTAL ACTION, c2020c; UEL, 201-?b).

Os dados transformados Z_{ij} são calculados pelo módulo da diferença entre cada observação original (X_{ij}) pela média original do nível i considerado ($\bar{X}_i$) (PORTAL ACTION, c2020c; UEL, 201-?b).

A hipótese H_0 sobre a igualdade das variâncias é rejeitada quando o valor de F^* é maior que F tabelado na linha $k-1$ e coluna $k(n-1)$, ou equivalentemente, quando a probabilidade do teste $p < \alpha$ (PORTAL ACTION, c2020c; UEL, 201-?b).

Além da análise gráfica e do teste de Levene, outros testes podem ser utilizados para verificar a homogeneidade de variâncias do conjunto de dados, como os testes de Bartlett (resumo geral disponível no apêndice A), Breusch-Pagan, Goldfeld-Quandt e de Cochran (PORTAL ACTION, c2020c).

1.1.4 Testes de hipóteses para comparação de grupos

Muitos estudos biológicos buscam verificar o efeito de tratamentos aplicados nos experimentos, ou o efeito de uma variável de interesse em amostra de uma população, e para este fim, utilizam de testes de hipóteses para a comparação de grupos (SOKAL; ROHLF, 1995; ZAR, 1999).

Quando os valores dos grupos a serem comparados apresentam distribuição normal e igualdade das variâncias entre si, pode-se utilizar de testes de hipóteses paramétricos, como o teste t (resumo geral disponível no apêndice B) e a ANOVA (*Analysis of Variance*), para comparação das médias de duas ou mais amostras. Quando os dados não atendem os pressupostos anteriores, deve-se utilizar de testes não-paramétricos para comparação de grupos, como por exemplo, o teste de Kruskal-Wallis e o teste de Friedman. Os testes não-paramétricos geralmente atribuem postos em ordem crescente ou decrescente às observações, não baseando suas análises em parâmetros, como a média, mediana e variância (AYRES, 2011; ZAR, 1999).

Caso ocorra a rejeição da hipótese nula (H_0) nos testes de comparação de grupos, a qual se refere existência de igualdade entre grupos, pode verificar quais destes diferem-se entre si por meio de testes de comparações múltiplas, como por exemplo, o teste de Tukey, teste de Student-Newman-Keuls e teste de Dunnett (AYRES, 2011; ZAR, 1999).

1.1.4.1 Análise de variância (ANOVA)

Em casos que os dados coletados apresentam replicação e randomização adequadas, independência, distribuição normal e homocedasticidade entre os grupos, pode-se utilizar os testes de ANOVA para comparação de igualdade de médias ($\mu_1 = \mu_2 = \dots \mu_i$) entre as amostras ou grupos ($i=1,2,\dots,a$) testando a hipótese nula destes derivarem de uma mesma população (GOTELLI; ELLISON, 2011; ZAR, 1999).

Os testes de ANOVA podem ser aplicados para avaliar a influência de um ou mais fatores (tratamentos) e de seus níveis nas médias dos grupos amostrais, apresentando variações em seus cálculos dependendo da natureza dos dados estudados (ANOVA tipo I, II ou III), número de fatores e variáveis (resultados) considerados. Seus testes se baseiam nas estimações das somas dos quadrados (total, dentro dos grupos, entre os grupos, dos fatores e de suas interações) para posterior cálculo dos quadrados médios e de suas respectivas razões F , as quais poderão rejeitar as hipóteses nulas de igualdade entre as médias ou ausência de efeito dos fatores, seus níveis e de suas interações, quando apresentarem valores de F maiores que aqueles tabelados para um determinado α do teste e grau de liberdade, ou equivalentemente, quando apresentar probabilidade do teste $p < \alpha$ (AYRES, 2011; GOTELLI; ELLISON, 2011; ZAR, 1999).

A ANOVA do tipo I ou modelo de efeitos fixos apresenta os cálculos mais simplificados, sendo que alguns são apresentados no apêndice C, no final do trabalho. Este modelo é utilizado em estudos cujos fatores e seus níveis foram determinados previamente. A ANOVA do tipo II ou modelo de efeitos aleatórios, os fatores e seus níveis são escolhidos aleatoriamente, geralmente após a coleta de dados. O modelo II apresenta algumas diferenças nos cálculos em relação ao modelo I, tornando as análises mais complexas ou inviáveis para estudos que consideram inúmeros fatores e variáveis. O modelo III ou misto, constitui-se de mescla de cálculos dos tipos anteriormente citados, decorrente de um dos fatores ser fixo, enquanto o outro escolhido aleatoriamente (SOKAL; ROHLF, 1995; ZAR, 1999).

1.1.4.1.1 ANOVA de blocos aleatorizados

A ANOVA “two way” ou blocos aleatorizados constitui-se de uma ANOVA de dois fatores, em que o primeiro fator constitui-se nos tratamentos ou grupos e seus níveis (A , com $i=1,2,...,a$ níveis), e o segundo fator os blocos (B , com $j=1,2,...,b$ níveis). Os blocos são estruturas caracterizadas por apresentar homogeneidade de condições em seu interior, e heterogeneidade entre os diferentes blocos, sendo que em cada um, os tratamentos são distribuídos aleatoriamente. Cada bloco poderia se constituir, por exemplo, indivíduos de uma mesma espécie, que recebem diferentes tipos de alimentos, os quais, por sua vez, seriam os níveis do fator (GOTELLI; ELLISON, 2011; ZAR, 1999).

Este tipo de delineamento se caracteriza por constituir de uma ANOVA de modelo III ou misto, em que o tratamento ou grupo constitui-se do fator fixo, enquanto os blocos corresponderiam ao fator aleatório (ZAR, 1999). Desta forma, a soma dos quadrados dentro dos tratamentos ou grupos ($SQ_{dentro\ de\ grupos}$), é calculada da seguinte forma:

$$SQ_{dentro\ de\ grupos} = \sum_{i=1}^a \sum_{j=1}^b (Y_{ij} - \bar{Y}_i - Y_j + \bar{Y})^2 \quad (1.05)$$

Sendo que Y_{ij} consiste da observação de um determinado tratamento ou grupo $i=1,2,...,a$ níveis e bloco $j=1,2,...,b$ níveis; $\bar{Y}_i$ a média do tratamento ou grupo; Y_j o bloco em consideração; e $\bar{Y}$ a média geral (GOTELLI; ELLISON, 2011).

Além disso, para este tipo de análise calculam-se também as somas de quadrados entre grupos ($SQ_{entre\ grupos}$) e dos blocos (SQ_{blocos}). A estimativa da $SQ_{entre\ grupos}$ é obtida pela seguinte formula:

$$SQ_{entre\ grupos} = \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_i - \bar{Y})^2 \quad (1.06)$$

A SQ_{blocos} é calculada através da equação a seguir:

$$SQ_{blocos} = \sum_{i=1}^a \sum_{j=1}^b (\bar{Y}_j - \bar{Y})^2 \quad (1.07)$$

Em que $\bar{Y}_j$ representa a média de um bloco $j=1,2,...,b$ níveis (GOTELLI; ELLISON, 2011).

Os cálculos dos quadrados médios entre grupos ($QM_{entre\ grupos}$), blocos (QM_{blocos}) e dentro dos grupos ($QM_{dentro\ dos\ grupos}$) ocorrem pela razão das somas dos quadrados pelos seus respectivos graus de liberdade $a-1$, $b-1$ e $(a-1)(b-1)$. Calculados os valores dos quadrados médios, obtêm-se duas razões F pela divisão de $QM_{entre\ grupos}/QM_{dentro\ dos\ grupos}$ e

$QM_{blocos}/QM_{dentro\ dos\ grupos}$. Geralmente a razão F obtida pelo QM_{blocos} não é de muito interesse do pesquisador, decorrente do intuito principal de averiguar o efeito dos tratamentos aplicados, somada a dificuldade de conhecer o possível efeito de interação entre os blocos. Desta forma, compara-se o valor da razão F obtida pelo $QM_{entre\ grupos}$ com o valor da tabela F para um determinado nível α do teste, numerador $a-1$, e denominador $(a-1)(b-1)$, ou equivalentemente, compara-se o valor da probabilidade do teste com o erro α escolhido (GOTELLI; ELLISON, 2011; ZAR, 1999).

1.1.4.2 Teste de Kruskal-Wallis

O teste de Kruskal-Wallis constitui-se de um teste não-paramétrico, comumente utilizado para comparar diferenças entre grupos de dados independentes $k > 2$, sendo aplicável para amostras de tamanhos desiguais, com variâncias heterogêneas e distribuição não normal dos dados. Neste teste, os dados recebem postos em ordem crescente dos valores das observações, os quais são posteriormente somados por grupos ($i=1, 2...k$) e comparados (AYRES, 2011; ZAR, 1999).

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (1.08)$$

Em que N consiste na quantidade total de observações; R_i e n_i a soma de postos e o número de observações em um grupo i , respectivamente. Quando o valor H calculado apresentar-se maior ou igual ao tabelado na tabela qui-quadrado para grau de liberdade $k-1$ e determinado nível α do teste, ou quando a probabilidade do teste constituir-se de $p < \alpha$, rejeita-se a hipótese nula, aceitando a existência de diferenças significativas entre as amostras (AYRES, 2011; ZAR, 1999).

1.1.4.3 Teste de Friedman

O teste de Friedman constitui-se de um teste não-paramétrico alternativo para o delineamento de blocos aleatorizados, permitindo a comparação de dois ou mais grupos não independentes ($i=1,2...a$ níveis), e não necessariamente normais ou homocedásticos, organizados em b blocos ($j=1,2,...b$ blocos), que possuam a mesma quantidade de réplicas (n) por combinação dos níveis dos grupos e blocos. Neste teste atribuem-se postos às observações de cada bloco, em ordem crescente de valores, realizando, posteriormente, a soma dos postos por grupo (R_i) (ZAR, 1999).

$$X_r^2 = \frac{12}{ban^2(na + 1)} \sum_{i=1}^a R_i^2 - 3b(na + 1) \quad (1.09)$$

Quando o X_r^2 resultante do teste apresenta valor menor que o tabelado na tabela qui-quadrado, referente ao grau de liberdade $a-1$ e determinado nível de erro α , ou quando a probabilidade do teste $p \geq \alpha$, se aceita a hipótese nula acerca da igualdade entre os grupos (AYRES, 2011; ZAR, 1999).

1.1.5 Correlação

A análise de correlação possibilita averiguar a existência de associação entre variáveis, as quais nem sempre apresentam relação de dependência. A intensidade da relação geralmente é representada pelo coeficiente de correlação amostral de Pearson (r) para uma amostra com distribuição normal bivariada (AYRES, 2011; ZAR, 1999).

1.1.5.1 Correlação de Spearman

Quando não ocorre distribuição bivariada normal em uma amostra, pode-se utilizar o coeficiente de correlação de Spearman. Este tipo de análise estima o coeficiente de correlação r_s através da atribuição de ranks em ordem crescente de valor para cada variável de uma amostra $i=1,2,...,n$. Desta forma, é possível posteriormente calcular as diferenças (d_i) para cada par de ranks (AYRES, 2011; ZAR, 1999). O coeficiente é dado por:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (1.10)$$

O coeficiente de Spearman varia de $-1 \leq r_s \leq +1$, no qual os sinais indicam a direção da correlação, positiva (valores de ambas variáveis elevam-se) ou negativa (uma variável aumenta e outra reduz), enquanto a proximidade com o valor absoluto 1 refere-se a intensidade da associação, sendo que $r_s=0$ indica a ausência de correlação (AYRES, 2011; ZAR, 1999).

Além disso, também é possível realizar testes de hipóteses sobre a significância de ρ (coeficiente de correlação para a população), rejeitando-se a hipótese nula de ausência de associação entre as variáveis quando $r_s > r_{s\ tab}$ (coluna α e linha n da tabela de correlação rankeada de Spearman) ou quando a probabilidade do teste $p \leq \alpha$ (AYRES, 2011; ZAR, 1999).

1.1.6 *Análise de sobrevivência*

A análise de sobrevivência consiste em estudos realizados por um extenso período de tempo, verificando principalmente a morbidade ou mortalidade de uma população, sendo utilizados comumente na área de saúde por possibilitarem a avaliação do efeito de medicamentos ou tratamentos na sobrevivência, ou no tempo de recuperação dos pacientes (COLOSIMO; GIOLO, 2006). Segundo Ayres (2011), esse tipo de pesquisa pode ser censitária ou amostral.

1.1.6.1 Estudos censitários

Os estudos censitários procuram realizar o censo sobre o tamanho de uma população, assim como de outras características de interesse da pesquisa, como a idade, sexo, profissão, sobrevivência, entre outros. Neste tipo de estudo, os dados geralmente são resumidos na forma de tabelas, gráficos ou pirâmides populacionais (AYRES, 2011).

1.1.6.2 Estudos amostrais

Os estudos amostrais de sobrevivência são realizados por meio da análise de uma ou mais amostras de uma população, podendo comparar grupos com mesmos ou diferentes tratamentos, ou estimar a probabilidade de sobrevivência. Este tipo de pesquisa diferencia dos outros testes estatísticos, pois considera indivíduos que ingressaram posteriormente ao início do trabalho ou abandonaram (“censurados” ou “progressivamente censurados”), assim como tem como principal variável resposta o tempo de falha ou de ocorrência do evento desejado (AYRES, 2011; COLOSIMO; GIOLO, 2006).

O tempo de falha é composto pelo tempo de início do estudo ou do diagnóstico, unidade de tempo e o evento de falha. Este último consiste no evento de interesse avaliado nos estudos de análise de sobrevivência, o qual pode consistir no falecimento, cura ou redução de sintomas de uma patologia, ou reaparecimento de enfermidade, que podem ser consequências da doença estudada ou do tratamento aplicado (COLOSIMO; GIOLO, 2006).

Alguns modelos comumente utilizados na análise de sobrevivência são: Atuarial, Kaplan-Meier, Nelson-Aalen, teste Log-Rank, e Modelo de risco proporcional de Cox (AYRES, 2011; COLOSIMO; GIOLO, 2006). Os modelos de Kaplan-Meier e do teste Log-

Rank são apresentados a seguir, enquanto os resumos dos modelos Atuarial, Nelson-Aalen e do risco proporcional de Cox, encontram-se no apêndice G.

1.1.6.2.1 Kaplan-Meier

O modelo não-paramétrico de Kaplan-Meier estima a probabilidade de sobrevivência de indivíduos de uma amostra, considerando os valores referentes ao número de indivíduos sobreviventes e as falhas ou ocorrências. O modelo de Kaplan-Meier não requer registro de informações em intervalos constantes, os quais são estimados pelos tempos de início e término do estudo, e pelos seus distintos tempos de ocorrência t_j , sendo $j=1,2,\dots,k$, e k os tempos de falha dessemelhantes em ordem crescente. Desta forma, este modelo possibilita a obtenção da proporção exata de sobrevivência (AYRES, 2011; COLOSIMO; GIOLO, 2006).

Quando não ocorrem censuras no conjunto de dados analisados, pode-se estimar a função de sobrevivência acumulada de Kaplan-Meier ($\hat{S}_{KM}(t)$) pela seguinte equação:

$$\hat{S}_{KM}(t) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right) \quad (1.11)$$

Sendo d_j as ocorrências no tempo t_j ; e n_j os sobreviventes ou número de indivíduos sob risco em t_{j-1} . Esse coeficiente de sobrevivência acumulada mede a probabilidade de um sujeito sobreviver até um tempo t_j , considerando que este sobreviveu aos tempos de falhas anteriores (CARVALHO; CAMPOS; OLIVEIRA, 200-?; COLOSIMO; GIOLO, 2006).

Além da função de sobrevivência acumulada, também é possível estimar a probabilidade de indivíduos presenciarem falhas ($\hat{q}_j$) no intervalo entre dois tempos de ocorrência distintos t_{j-1} e t_j , considerando que estes sobreviveram aos eventos anteriores:

$$\hat{q}_j = \frac{d_j}{n_j} \quad (1.12)$$

Outra função comumente utilizada na análise de sobrevivência consiste na taxa de falha ($\hat{\lambda}(t)$), a qual expressa a quantidade de eventos que ocorrem em uma população por uma unidade de tempo (CARVALHO; CAMPOS; OLIVEIRA, 200-?; COLOSIMO; GIOLO, 2006).

$$\hat{\lambda}(t) = \frac{d_j}{n_j (\Delta_t)} \quad (1.13)$$

Sendo que Δ_t representa o tamanho do intervalo entre dois tempos distintos de falhas consecutivos t_{j-1} e t_j (CARVALHO; CAMPOS; OLIVEIRA, 200-?; COLOSIMO; GIOLO, 2006).

1.1.6.2.2 Log-Rank test

O teste logrank constitui-se de teste não-paramétrico que possibilita a comparação de duas ou mais funções de sobrevivência ($S_1(t)$ e $S_2(t)$), verificando sua igualdade, por meio das diferenças entre o número de perdas “esperadas” e “observadas” das amostras ou grupos i considerados ($i=1,...,r$), as quais podem ser calculadas pelos valores de número de indivíduos em risco (n_{ij}) e de ocorrências (d_{ij}) em tempos de falhas distintos t_j ($j=1,...,k$) (AYRES, 2011; COLOSIMO; GIOLO, 2006).

O número de perdas esperadas de um grupo (E_{ij}) para um determinado tempo de falha constitui-se do produto entre o total de ocorrências em t_j ($\sum_{i=1}^r d_{ij}$), pela razão de sobreviventes em t_{j-1} no grupo considerado (n_{ij}), em relação ao total de pacientes sob risco em t_{j-1} ($\sum_{i=1}^r n_{ij}$). Os cálculos apresentados neste parágrafo estão representados mais claramente na seguinte equação:

$$E_{ij} = \sum_{i=1}^r d_{ij} \frac{(n_{ij})}{(\sum_{i=1}^r n_{ij})} \quad (1.14)$$

Após o cálculo de cada valor de perdas esperadas para todas as amostras e tempos de ocorrências considerados, realiza-se a soma destes, e posteriormente calcula-se a diferença entre o total de perdas esperadas e observadas.

$$X^2 = \sum_{i=1}^r \left[\frac{(|\sum_{j=1}^k O_{ij} - \sum_{j=1}^k E_{ij}| - 0,5)^2}{\sum_{j=1}^k E_{ij}} \right] \quad (1.15)$$

Sendo O_{ij} o número de perdas observadas para uma determinada amostra ou grupo i e tempo t_j (AYRES, 2011).

Caso a probabilidade do teste seja $p > \alpha$ ou o valor $X^2 \leq X^2_{tab}$ tabelado para a linha $k-1$ e coluna α , na tabela qui-quadrado, se aceita a hipótese nula sobre a igualdade de sobrevivência das amostras comparadas, indicando, por exemplo, que medicamentos ou tratamentos aplicados apresentam efeitos semelhantes nos grupos (AYRES, 2011; COLOSIMO; GIOLO, 2006).

1.2 Dípteros como modelo biológico

Considerando que dois dos três conjuntos de dados biológicos que foram analisados neste trabalho se referem a dípteros, realizou-se uma pesquisa sobre alguns aspectos biológicos que auxiliariam na interpretação dos resultados.

Os dípteros (Classe: Insecta, Filo: Arthropoda) são constantemente objetos de estudos biológicos e ecológicos, decorrente de algumas espécies serem polinizadoras, outras auxiliam na reciclagem de matéria orgânica, enquanto outras causam prejuízos à saúde humana, animal e vegetal, por constituírem de vetores biológicos de doenças, pragas em plantações ou ectoparasitos. Estes insetos são comumente utilizados como modelos biológicos para pesquisas em genética, evolução, ecologia e biologia da conservação devido às dimensões e ciclo de vida reduzidos, favorecendo o estudo do desenvolvimento de dípteros ao longo de várias gerações (BONIN, 2010).

1.2.1 Moscas-das-frutas

Entre os dípteros, algumas espécies de drosofilídeos vêm causando enormes prejuízos na economia agroexportadora brasileira, devido à realização da oviposição no ostíolo do figo-roxo (*Ficus carica*) antes do seu período de maturação, o qual será colonizado e consumido pelas larvas do inseto, tornando-se impróprio para o consumo humano (BOAVENTURA, 2013). Moscas-das-frutas, como a espécie invasora *Zaprionus indianus* (Grupta), oriunda da região Afrotropical, causou perda de 50 à 80% da produção da variedade de figo-roxo de Valinhos entre os anos de 2000 e 2001 (BOAVENTURA, 2013; MATAVELLI, 2014).

As espécies invasoras constituem de organismos que ocorrem naturalmente em outros países ou continentes, e que quando introduzidas acidentalmente ou propositalmente pela ação humana em um novo espaço, passam a reproduzir, dispersar e competir por alimento e espaço com espécies nativas (ZILLER; ZALBA, 2007). A competição interespecífica entre espécies invasoras e nativas pode ocasionar alterações nos ecossistemas, e consequentemente, reduzir ou até extinguir organismos nativos, culminando na diminuição da biodiversidade local (ROHDE *et al.*, 2010).

Considerando os problemas econômicos e ambientais envolvendo moscas-das-frutas, Matavelli (2014), Bonin (2010) e Boaventura (2013) ressaltam a importância de estudar aspectos biológicos sobre a flutuação de suas populações no decorrer do ano e fatores que influenciam sua abundância. De acordo com Bonin (2010) e Silva (2013), a abundância de drosofilídeos apresenta-se associada ao seu desenvolvimento, o qual pode ser favorecido pela disponibilidade de substrato em grande quantidade, condição razoável de temperatura (intervalo tolerável de 10 °C e 38 °C, e temperatura ótima de 25 °C) e elevada umidade.

Desta forma, os drosofilídeos geralmente seriam mais abundantes em estações mais quentes e chuvosas, que se estenderiam de novembro a maio no clima subtropical úmido de

inverno seco e verão quente, pelas altas temperaturas e umidade propiciarem a frutificação da figueira, assim como a redução da duração dos estágios de desenvolvimento das moscas (ALVARES *et al.*, 2014; BONIN, 2010; LEONEL; SAMPAIO, 2011; ZAMPAULO *et al.*, 2007). Segundo os estudos de Tantawy & Vetukhiv (1960) e Tantawy (1961) acerca da espécie *Drosophila pseudoobscura*, quando o desenvolvimento destes insetos é favorecido, comumente observa-se o aumento do seu tamanho corpóreo, longevidade e da quantidade de ovos depositados por fêmeas, consequentemente, elevando o número de indivíduos que compõem a população (BONIN, 2010).

Além dos fatores discutidos anteriormente, a abundância de drosofilídeos também é influenciada por características espécies-específicas, relacionadas ao nicho ecológico da espécie, conceito que, segundo Odum e Barrett (2008), refere-se à função desempenhada por uma espécie em uma comunidade (posição trófica na teia alimentar, ocupação, taxa de crescimento, aptidão, capacidade de adaptação, interações com outros seres e com o ecossistema), englobando também o seu habitat, suas necessidades e exigências ambientais (temperatura, pH, umidade, recursos alimentares) (MATAVELLI, 2014; TORRES; RAVAZZI, 2006). Considerando as características intrínsecas das espécies, aquelas como hábitos alimentares generalistas e alta capacidade de adaptação às condições adversas, atributos geralmente presentes em espécies invasoras, apresentariam maior sucesso na ocupação do ambiente (ROHDE *et al.*, 2010).

1.2.2 *Chrysomya megacephala*

Diferentemente das moscas-das-frutas, a espécie exótica *Chrysomya megacephala* (Fabricius), constitui-se de um tipo de mosca-varejeira (Família: Calliphoridae) originária da Australásia, que foi introduzida em terras americanas por navios vindos da África (BAUMGARTNER; GREENBERG, 1984). O gênero *Chrysomya*, apresenta grande importância médica, veterinária e sanitária pelas espécies serem causadoras de miíases em seres humanos e outros animais, assim como são responsáveis pela disseminação de diversas doenças virais, bacterianas e verminoses (GOMES; VON ZUBEN; SANCHES, 2003). Além disso, a espécie *C. megacephala* é muito utilizada na entomologia forense, por ser indicadora do tempo de deterioração de cadáveres humanos, devido às características de oviposição e alimentação a partir de substratos efêmeros (VON ZUBEN *et al.* 1996; WELLS & GREENBERG 1992).

A espécie *C. megacephala* caracteriza-se por apresentar oviposição agregada, decorrente da ação de feromônios liberados no processo que atraem outras fêmeas para colocar seus ovos no mesmo substrato (NORRIS 1965; VON ZUBEN; STANGENHAUS; GODOY, 2000). Ao eclodir, as larvas passam a se alimentar do substrato até consumi-lo totalmente, e posteriormente se dispersam para empupar, caso tenham atingido um peso mínimo, ou passam a procurar por mais alimento (GOMES; VON ZUBEN; SANCHES, 2003). Segundo estudos de Von Zuben *et al.* (1993), quando ocorre alta densidade larval em um mesmo substrato, decorrente da menor quantidade de recursos per capita, sucede-se a competição intraespecífica por alimento, a qual geralmente culmina em uma redução da taxa de sobrevivência da maioria das larvas, diminuição das dimensões do adulto decorrente do atraso no desenvolvimento, assim como uma redução da fecundidade.

Considerando a relevância dos dípteros para polinização, decomposição, reciclagem de carbono na natureza, assim como o impacto de suas populações na economia agroexportadora e nas áreas médica, veterinária e sanitária, estudos sobre o desenvolvimento, variação populacional e ecologia de dípteros são muito importantes, podendo servir como base para pesquisas que visem o controle de suas populações ou que auxiliem na interpretação e identificação do tempo de decomposição de cadáveres, essenciais para processos criminais (BONIN, 2010; RAGA; FILHO; 2003; VON ZUBEN *et al.* 1996; WELLS; GREENBERG 1992).

1.3 Objetivos

Este trabalho apresentou como objetivo o aprendizado teórico sobre análise exploratória de dados, testes de normalidade e homocedasticidade, testes paramétricos e não paramétricos para comparações de grupos, transformações de dados, análise de correlação e regressão, análise de sobrevivência e análise de componentes principais. Além disso, aplicaram-se alguns dos testes estatísticos estudados na análise de conjunto de dados biológicos, concomitante ao aprendizado acerca do funcionamento do programa R 3.6.1 (R CORE TEAM, 2019). A teoria estudada foi aplicada a três conjuntos de dados: moscas coletadas em uma cultura de figo-roxo no decorrer de um ano; posturas de moscas da espécie *Chrysomya megacephala*, em um experimento de laboratório; dados necrológicos do município de Bebedouro, SP.

2 MATERIAIS E MÉTODOS

Neste trabalho foi realizado o estudo teórico sobre análise exploratória de dados, testes de normalidade e homocedasticidade, testes paramétricos e não paramétricos para comparações de grupos, transformações de dados, análise de sobrevivência e análise de componentes principais por meio da leitura de livros de Zar (1999), Gotelli, Ellison (2011), Ayres (2011), Sokal, Rohlf (2012), Colosimo, Giolo (2006), Manly (2008) e Mingoti (2005).

Para a aplicação de alguns dos testes estatísticos estudados, foram utilizados conjuntos de dados biológicos fornecidos pelo Prof. Dr. Cláudio Von Zuben, acerca da contagem de espécimes de moscas-das-frutas de diferentes espécies, coletadas em culturas de figo-roxo no município de Valinhos, SP, durante o período de setembro de 2010 à agosto de 2011, assim como dados referentes à postura de moscas da espécie *Chrysomya megacephala*, em laboratório, submetidas a ambientes com diferentes densidades de califorídeos adultos (100, 200, 400 ou 800 dípteros por gaiola ou réplica), e sob diferentes posicionamentos do material substrato. Além destes conjuntos de dados, aplicou-se os conhecimentos de análise de sobrevivência em dados necrológicos humanos de 1 de janeiro até 31 de dezembro de 2019 da cidade de Bebedouro, SP, sendo apenas consideradas as variáveis disponíveis sobre a idade e o sexo dos indivíduos, não havendo acesso aos nomes dos falecidos.

Desta forma, agruparam-se as moscas-das-frutas em 8 gêneros, sendo que apenas 4 foram considerados na análise por apresentarem número de coletas anuais superior a 100 indivíduos. Para este conjunto, considerando um delineamento de blocos aleatorizados, foram aplicados testes de comparações paramétricos e não-paramétricos para verificar a existência de diferenças significativas dos grupos (meses ou estações), possivelmente decorrente das mudanças das condições ambientais durante o ano (temperatura, umidade e disponibilidade de substrato), considerando o efeito dos blocos (gêneros) (BONIN, 2010).

Para a comparação entre os meses, utilizou-se do teste de Friedman. Em uma segunda análise, buscando verificar a influência das estações na abundância de drosofilídeos, aplicou-se o teste ANOVA de blocos aleatorizados nos dados referentes aos meses agrupados em estações: primavera (outubro, novembro e dezembro); verão (janeiro, fevereiro e março); outono (abril, maio e junho) e inverno (julho, agosto e setembro) (CARRER; GARCIA, 2007; NOVAIS; BRITO, 2017). Antes da execução do teste ANOVA foram verificados os seus pressupostos sobre normalidade e homocedasticidade dos dados pela aplicação dos testes Shapiro-Wilk e Levene, respectivamente, assim como a análise gráfica dos resíduos. Nas comparações múltiplas da ANOVA de blocos aleatorizados, utilizou o teste de Tukey.

Para este primeiro conjunto de dados também foi realizada a análise de correlação de Spearman entre gêneros para avaliar a possível existência de associações negativas, que poderiam sugerir situações de competições interespecíficas (ROHDE *et al.*, 2010). Além disso, realizou-se análise gráfica das frequências relativas e absolutas dos gêneros considerados, buscando observar a flutuação da abundância dos drosofilídeos ao decorrer do ano (BONIN, 2010).

Na análise do segundo conjunto de dados, procurando verificar a ocorrência aleatória de oviposição agregada de *C. megacephala* em substratos nos interiores de potes, os quais, por sua vez, estavam contidos dentro de gaiolas (réplicas) com determinada densidade de moscas (100, 200, 400 e 800 califorídeos por gaiola), aplicou-se o teste de Kruskal-Wallis para comparar a quantidade de posturas nos quatro potes (grupos) dentro de cada uma das réplicas (NORRIS 1965; VON ZUBEN; STANGENHAUS; GODOY, 2000). Para as comparações múltiplas foi utilizado o teste de Student-Newman-Keuls.

No terceiro conjunto de dados, calcularam-se o estimador de sobrevivência acumulada de Kaplan-Meier para cada idade de falecimento, assim como probabilidades e taxas de falhas, construindo-se representações gráficas e histogramas para as diferentes idades e sexos, buscando avaliar a ocorrência de diferenças entre as curvas de sobrevivência para a população feminina e masculina do município de Bebedouro, SP. Em que a probabilidade de falha consiste na possibilidade de ocorrência de falecimentos em determinada idade, enquanto a taxa de falha expressa a quantidade de óbitos por ano, considerando o intervalo entre tempos distintos de falha consecutivos. Posteriormente utilizou o teste Log-rank para averiguar se as diferenças entre as curvas de sobrevivência dos sexos eram significativas.

Os testes de normalidade, homocedasticidade, comparações de grupos, múltiplas e correlação foram realizados considerando o nível de erro $\alpha=0,05$. As análises estatísticas foram realizadas nos programas BioEstat 5.3, Microsoft Office Excel 2007 e R 3.6.1. (AYRES *et al.* 2007; MICROSOFT, 2007; R CORE TEAM, 2019). No programa R utilizou os pacotes stats, car, moments e survival (FOX; WEISBERG, 2019; KOMSTA; NOVOMESTKY, 2015; R CORE TEAM, 2019; THERNEAU, 2015).

3 RESULTADOS E DISCUSSÃO

3.1 Análise do conjunto de dados de moscas-das-frutas agrupadas em gêneros

Nesta seção, apresentamos os resultados e a discussão referentes às análises dos dados das moscas-das-frutas agrupadas em gêneros com número de coletas superior a 100 indivíduos no período considerado (Tabela 1).

Tabela 1- Quatro gêneros de moscas-das-frutas com os maiores números de coletas durante o ano.

Gênero	set.	out.	nov.	dez.	jan.	fev.	mar.	abr.	maio	jun.	jul.	ago.	Total de indivíduos
<i>Zaprionus</i>	8	109	2675	21081	32476	7170	6840	9378	1918	1531	60	93	83339
<i>Drosophila</i>	291	2026	2624	2691	7438	1253	2943	3788	1554	1007	923	2134	28672
<i>Scaptodrosophila</i>	1	2	22	387	1188	860	4384	2141	3488	762	31	11	13277
<i>Rhinoleucophenga</i>	4	7	11	5	22	12	4	4	7	28	4	26	134

Fonte: Dados fornecidos pelo Prof. C. J. Von Zuben.

3.1.1 Comparação do número de coletas de moscas-das-frutas entre os meses

Inicialmente, antes de proceder com a comparação entre os meses, realizou-se uma breve análise exploratória das quantidades de coletas por gêneros, considerando todos os 12 meses. Primeiramente percebeu-se que, em geral, os dados apresentaram assimetria positiva ($g_1 > 0$) e curtose do tipo leptocúrtica ($g_2 > 3$), exceto para *Rhinoleucophenga* (Hendel), a qual a curtose era do tipo platicúrtica ($g_2 < 3$), na qual a densidade de probabilidade dos valores centrais é mais elevada em relação aos valores extremos, quando comparada a curtose do tipo leptocúrtica (Tabela 2).

Tabela 2- Medidas de posição e dispersão das quantidades de coletas por gêneros (*Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*), considerando todos os 12 meses.

Parâmetros	<i>Zaprionus</i>	<i>Drosophila</i>	<i>Scaptodrosophila</i>	<i>Rhinoleucophenga</i>
Mínimo	8,00	291,00	1,00	4,00
Primeiro Quartil (25%)	105,00	1191,50	19,25	4,00
Mediana	2296,50	2080,00	574,50	7,00
Terceiro Quartil (75%)	7722,00	2754,00	1426,25	14,50
Máximo	32476,00	7438,00	4384,00	28,00
Média Aritmética	6944,92	2389,33	1106,42	11,17
Variância	101787328,63	3504376,79	2196592,26	81,79
Desvio Padrão	10088,97	1872,00	1482,09	9,04
Coefficiente de Variação	145,27%	78,35%	133,95%	80,99%
Assimetria	1,65	1,66	1,25	0,96
Curtose	4,55	5,52	3,21	2,30

Fonte: dados da análise

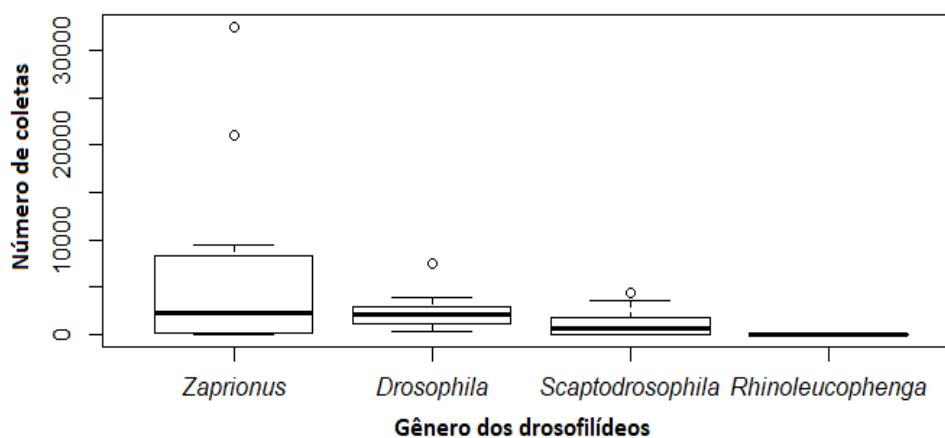
O gênero *Zaprionus* (Coquillett) apresentou o maior valor de total de coletas anuais, mesmo possuindo uma única espécie, como pode ser visualizada nas tabelas 1 e 3. Além disso, observou-se que o *Zaprionus* apresentou a maior média ($\bar{X}=6944,92$) e maior dispersão dos dados, considerando a variância ($s^2=101787328,63$), desvio padrão ($s=10088,97$) e o coeficiente de variação ($CV=145,27\%$) dentre os gêneros considerados (Tabela 2). Este fato ocorreu decorrente da presença de valores discrepantes (32476,00 e 21081,00) representados por círculos vazios no gráfico da figura 1, os quais elevaram o valor da média em relação à mediana ($Md=2296,50$), sendo responsável pela distribuição assimétrica dos dados a direita ($g_1=1,65$), como verificado na tabela 2.

Tabela 3- Quantidade de espécies de moscas-das-frutas agrupadas nos gêneros considerados.

Gênero	Número de espécies
<i>Zaprionus</i>	1
<i>Drosophila</i>	51
<i>Scaptodrosophila</i>	1
<i>Rhinoleucophenga</i>	5

Fonte: dados da análise

Figura 1- *Boxplot* das quantidades de coletas por gêneros (*Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*), considerando todos os 12 meses.



Fonte: dados da análise

Semelhantemente, para o gênero *Scaptodrosophila* (Duda) a assimetria à direita ($g_1=1,25$) foi ocasionada principalmente pela observação discrepante 4384 (Figura 1) que afastou a média ($\bar{X}=1106,42$) da mediana ($Md=574,50$). Por outro lado, mesmo com uma assimetria à direita ($g_1=1,66$) e medida discrepante 7438 (Figura 1), o gênero *Drosophila* (Fallén), composto pela maioria das espécies amostradas (51 spp), apresentou mediana ($Md=2080,00$) e média relativamente próximas ($\bar{X}=2389,33$). Esta questão pode ser decorrente de observações mais uniformes durante o ano quando comparado com outros gêneros (Tabela 1), como verificado pelo elevado valor do primeiro quartil de 1191,50 (Tabela 2), o qual divide os dados em dois, existindo 25% de valores inferiores a este limiar, e 75% superiores.

No gráfico da figura 1, as observações do gênero *Rhinoleucophenga* foram mascaradas pelo fator escala da figura, devido aos altos valores de coletas dos outros grupos, as quais variaram entre 1 à 32476, como pode ser verificado na tabela 1. Os dados para este gênero variaram entre 4 à 28, alcançando o valor máximo no mês de junho (Tabelas 1 e 2).

Quando analisou-se as quantidades de coletas por meses (setembro à agosto), considerando todos os gêneros, verificou-se, em geral, curtose do tipo platicúrtica ($g_2<3$), e assimetria a direita, exceto os meses de novembro ($\bar{X}=1333,00$; $Md=1323,00$), março ($\bar{X}=3542,75$; $Md=3663,50$), maio ($\bar{X}=1741,75$; $Md=1736,00$) e junho ($\bar{X}=832,00$; $Md=884,50$) que apresentaram maior simetria nas suas distribuições ($g_1\approx 0$), como pode-se verificar nas tabelas 4 e 5, assim como no gráfico da figura 2. Este gráfico também permitiu averiguar meses com altas variâncias e médias, decorrente de algumas observações elevadas, como verificado principalmente nos meses de dezembro ($\bar{X}=6041,00$; $Md=1539,00$) e janeiro

($\bar{X}$ =10281,00; M_d =4313,00), em detrimento as médias baixas de setembro ($\bar{X}$ =76,00; M_d =6,00) e julho ($\bar{X}$ =254,50; M_d =45,50), como observado nas tabelas 4 e 5. Desta forma, o teste de comparações de grupos avaliou se as diferenças entre esses meses seriam consideradas significativas.

Tabela 4- Medidas de posição e dispersão das quantidades de coletas por meses (setembro à fevereiro), considerando todos os 4 gêneros.

Parâmetros	set.	out.	nov.	dez.	jan.	fev.
Mínimo	1,00	2,00	11,00	5,00	22,00	12,00
Primeiro Quartil (25%)	3,25	5,75	19,25	291,50	896,50	648,00
Mediana	6,00	58,00	1323,00	1539,00	4313,00	1056,50
Terceiro Quartil (75%)	78,75	588,25	2636,75	7288,50	13697,50	2732,25
Máximo	291,00	2026,00	2675,00	21081,00	32476,00	7170,00
Média Aritmética	76,00	536,00	1333,00	6041,00	10281,00	2323,75
Variância	20552,67	989142,00	2311350,00	101941704,00	229543468,00	10706465,58
Desvio Padrão	143,36	994,56	1520,31	10096,62	15150,69	3272,07
Coefficiente de Variação	188,63%	185,55%	114,05%	167,13%	147,37%	140,81%
Assimetria	1,15	1,15	$5,4 \cdot 10^{-4}$	1,11	1,00	1,07
Curtose	2,33	2,33	1,00	2,29	2,20	2,27

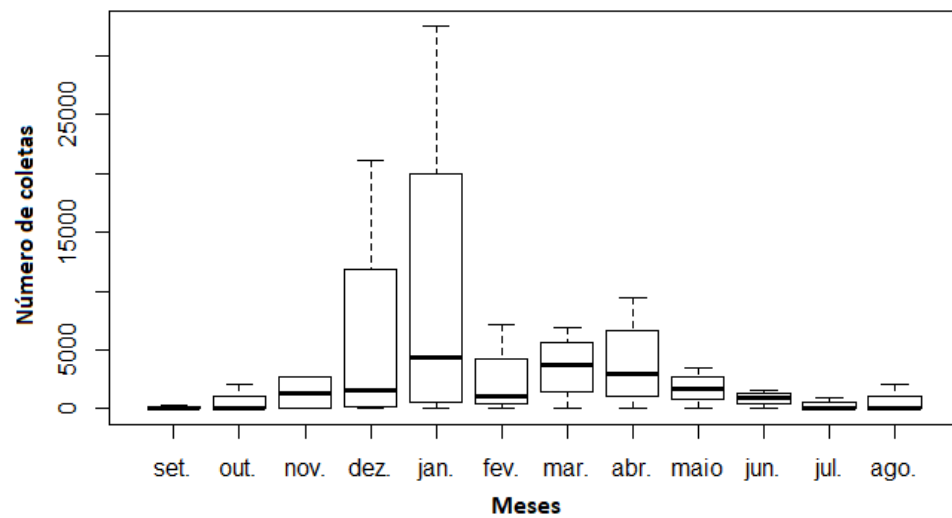
Fonte: dados da análise

Tabela 5- Medidas de posição e dispersão das quantidades de coletas por meses (março à agosto), considerando todos os 4 gêneros.

Parâmetros	mar.	abr.	maio	jun.	jul.	ago.
Mínimo	4,00	4,00	7,00	28,00	4,00	11,00
Primeiro Quartil (25%)	2208,25	1606,75	1167,25	578,50	24,25	22,25
Mediana	3663,50	2964,50	1736,00	884,50	45,50	59,50
Terceiro Quartil (75%)	4998,00	5185,50	2310,50	1138,00	275,75	603,25
Máximo	6840,00	9378,00	3488,00	1531,00	923,00	2134,00
Média Aritmética	3542,75	3827,75	1741,75	832,00	254,50	566,00
Variância	8154003,58	16091014,92	2041686,92	390180,67	199141,67	1093992,67
Desvio Padrão	2855,52	4011,36	1428,88	624,64	446,25	1045,94
Coefficiente de Variação	80,60%	104,80%	82,04%	75,08%	175,34%	184,80%
Assimetria	-0,13	0,66	0,014	-0,27	1,14	1,15
Curtose	1,84	2,01	1,96	1,92	2,33	2,33

Fonte: dados da análise

Figura 2- *Boxplot* das quantidades de coletas por meses (setembro à agosto), considerando todos os 4 gêneros.

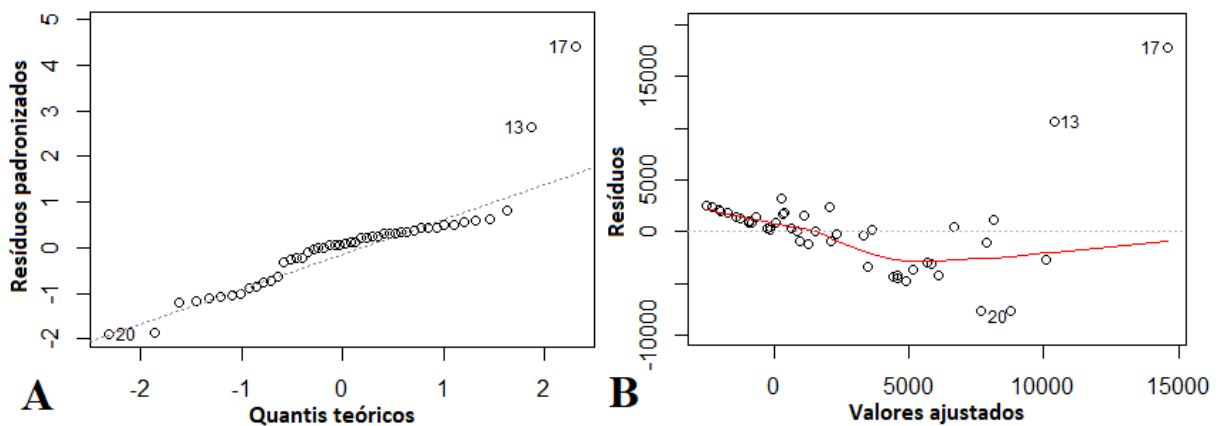


Fonte: dados da análise

Antes da aplicação do teste de comparações de grupos, verificou-se a normalidade e homocedasticidade das quantidades de coletas por meses, os pressupostos do teste ANOVA, considerando os quatro gêneros apresentados anteriormente.

Quando se aplicou o teste de Shapiro-Wilk aos resíduos dos dados, rejeitou-se a hipótese nula sobre a distribuição normal dos dados, decorrente dos valores de $W=0,816$ e $p=3,12 \cdot 10^{-6}$, sendo o último muito menor que $\alpha=0,05$. A análise do *qqplot* (Figura 3A) suportou o resultado deste teste, visto que os quantis dos resíduos apresentaram desvios da relação linear com os seus quantis teóricos.

Figura 3- **A.** Gráfico *qqplot*; **B.** Gráfico de resíduos em relação aos valores ajustados.



Fonte: dados da análise

Quanto ao teste de Levene, obteve-se, para 11 graus de liberdade, um valor $F^*=0,916$ e $p=0,535$, culminando na aceitação de H_0 sobre a homocedasticidade dos resíduos. Por outro lado, o gráfico da figura 3B, aparentou rejeitar a igualdade nas variâncias dos resíduos, visto que os pontos possuíam tendências na sua distribuição.

A partir dos resultados dos testes de normalidade e homocedasticidade, utilizou-se o teste de Friedman para comparação da quantidade de coletas entre os meses. O teste de Friedman, considerando 11 graus de liberdade, obteve o valor $X_r^2=21,683$ e $p=0,027$, rejeitando-se desta forma a hipótese nula sobre igualdade dos meses. O teste de comparações múltiplas revelou que apenas os meses de setembro e janeiro, que correspondiam à menor e a maior soma de ranks respectivamente (5,50 e 43,00), se diferenciaram significativamente ($p<0,05$) por uma diferença de 37,50.

A diferença significativa entre os meses de setembro e janeiro, verificada pelo teste de Friedman pode ser decorrente das diferenças de temperatura e umidade destes meses. Os meses de setembro e janeiro corresponderiam respectivamente ao inverno e verão, considerando a divisão do ano em quatro estações, e a estação seca e quente do clima subtropical úmido de inverno seco e verão quente, característico da região de Valinhos, SP, onde ocorreu a coletas de drosofilídeos (ALVARES *et al.*, 2014; CARRER; GARCIA, 2007; NOVAIS; BRITO, 2017; ZAMPAULO *et al.*, 2007). As temperaturas e umidade mais altas do verão estimulam a frutificação do figo, assim como a reprodução e o desenvolvimento das moscas, fator que explicaria o elevado número de coletas no mês de janeiro, e valor reduzido no mês de setembro (Tabela 1), os quais refletiram no valor final da soma dos ranks (BONIN, 2010; LEONEL; SAMPAIO, 2011).

3.1.2 Comparação do número de coletas de moscas-das-frutas entre estações

Os dados dos gêneros de moscas-das-frutas agrupados por estação são apresentados na tabela a seguir:

Tabela 6- Dados de moscas-das-frutas por gêneros agrupados em estações.

Gênero	Primavera	Verão	Outono	Inverno
<i>Zaprionus</i>	23865	46486	12827	161
<i>Drosophila</i>	7341	11634	6349	3348
<i>Scaptodrosophila</i>	411	6432	6391	43
<i>Rhinoleucophenga</i>	23	38	39	34

Fonte: dados da análise

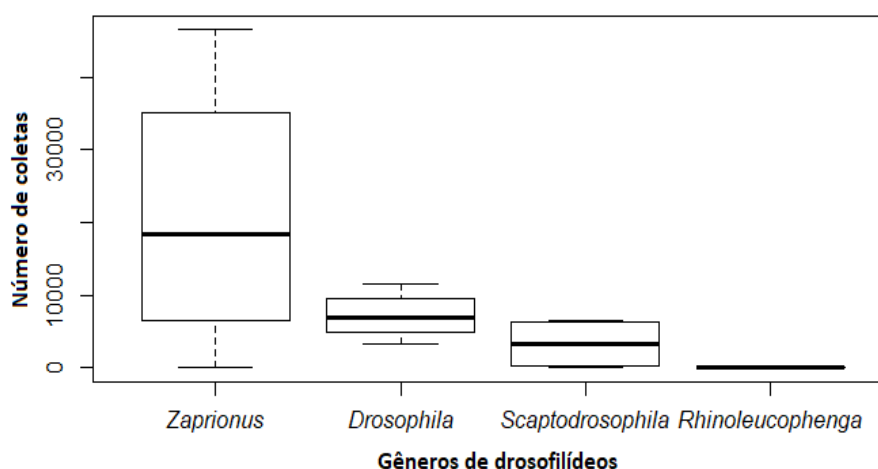
Uma análise exploratória inicial das quantidades de coletas por gêneros, considerando os dados agrupados em estações, permitiu observar que *Zaprionus* continuou apresentando os maiores valores para média ($\bar{X}=20834,75$), mediana ($Md=18346,00$), variância ($s^2=386232346,92$) e desvio padrão ($s=19652,79$) como podem ser vistos na tabela 7. Diferentemente da análise anterior, o maior coeficiente de variação pertenceu a *Scaptodrosophila* ($CV=107,67\%$).

Tabela 7- Medidas de posição e dispersão das quantidades de coletas por gêneros (*Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*), referentes aos dados agrupados em estações.

Parâmetros	<i>Zaprionus</i>	<i>Drosophila</i>	<i>Scaptodrosophila</i>	<i>Rhinoleucophenga</i>
Mínimo	161,00	3348,00	43,00	23,00
Primeiro Quartil (25%)	9660,50	5598,75	319,00	31,25
Mediana	18346,00	6845,00	3401,00	36,00
Terceiro Quartil (75%)	29520,25	8414,25	6401,25	38,25
Máximo	46486,00	11634,00	6432,00	39,00
Média Aritmética	20834,75	7168,00	3319,25	33,50
Variância	386232346,92	11746082,00	12772197,58	53,67
Desvio Padrão	19652,79	3427,26	3573,82	7,33
Coefficiente de Variação	94,33%	47,81%	107,67%	21,87%
Assimetria	0,38	0,31	-0,005	-0,88
Curtose	1,85	1,97	1,01	2,08

Fonte: dados da análise

Figura 4- Boxplot das quantidades de coletas por gêneros (*Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*), referentes aos dados agrupados em estações.



Fonte: dados da análise

A partir do gráfico da figura 4, percebeu-se que o agrupamento em estações favoreceu uma distribuição mais simétrica dos dados nos gêneros, os quais apresentaram médias e medianas próximas (Tabela 7), assim como evitou a presença de dados discrepantes, diferentemente da análise da seção 3.1.1. Estas mudanças no comportamento dos dados foram decorrentes das somas das observações de meses com elevados números de coletas com os de menores, e vice-versa, equilibrando os valores (como por exemplo, os meses de janeiro com fevereiro e março). Mesmo com agrupamento dos dados, as observações de *Rhinoleucophenga* foram mascaradas pelas altas quantidades de indivíduos coletados dos outros gêneros, e passaram a apresentar assimetria negativa ($g_2=-0,88$), com mediana ($Md=36,00$) maior que média ($\bar{X}=33,50$).

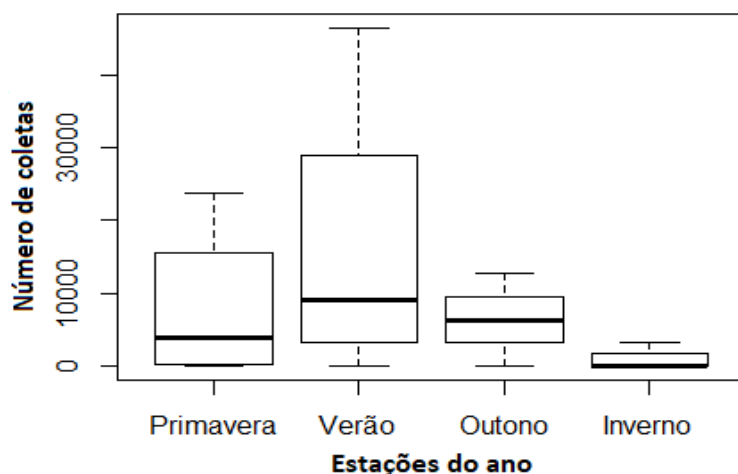
A análise exploratória das quantidades de coletas por estações, considerando todos os gêneros, permitiu verificar que estas apresentaram curtose do tipo platicúrtica ($g_2<3$), e assimetria positiva dos dados ($g_1>0$), exceto a estação de outono, a qual possuiu distribuição muito próxima da simétrica ($g_1\cong 0$), como pode ser observado na tabela 8 e figura 5.

Tabela 8- Medidas de posição e dispersão das quantidades de coletas por estações (primavera, verão, outono e inverno), considerando todos os gêneros.

Parâmetros	Primavera	Verão	Outono	Inverno
Mínimo	23,00	38,00	39,00	34,00
Primeiro Quartil (25%)	314,00	4833,50	4771,50	40,75
Mediana	3876,00	9033,00	6370,00	102,00
Terceiro Quartil (75%)	11472,00	20347,00	8000,00	957,75
Máximo	23865,00	46486,00	12827,00	3348,00
Média Aritmética	7910,00	16147,50	6401,50	896,50
Variância	124441852,00	431567731,67	27257107,67	2674393,67
Desvio Padrão	11155,35	20774,21	5220,83	1635,36
Coefficiente de Variação	141,03%	128,65%	81,56%	182,42%
Assimetria	0,87	0,98	0,02	1,15
Curtose	2,06	2,20	2,00	2,33

Fonte: dados da análise

Figura 5- *Boxplot* das quantidades de coletas por estações (primavera, verão, outono e inverno), considerando todos os gêneros.



Fonte: dados da análise

Observou-se também que os maiores valores de média ($\bar{X}=16147,50$), mediana ($Md=9033,00$), variância ($s^2=431567731,67$) e desvio padrão ($s=20774,21$) ocorreram na estação de verão, assim como uma assimetria à direita ($g_1=0,98$), devido às observações de números altos dos gêneros *Zaprionus* e *Drosophila* no mês de janeiro (32476 e 7438, respectivamente, como mostrada na tabela 1), principalmente, as quais elevaram a média desta estação.

Após o verão, a primavera apresentou as maiores medidas de média ($\bar{X}=7910,00$), variância ($s^2=124441852,00$) e desvio padrão ($s=11155,35$), sendo que a assimetria à direita causada pelos altos valores de coletas para *Zaprionus* durante o mês de dezembro (21081), como verificado na tabela 1. A estação de outono possuiu média ($\bar{X}=6401,50$) e mediana ($Md=6370,00$) próximas, mostrando uma distribuição de dados mais uniforme (Tabela 8).

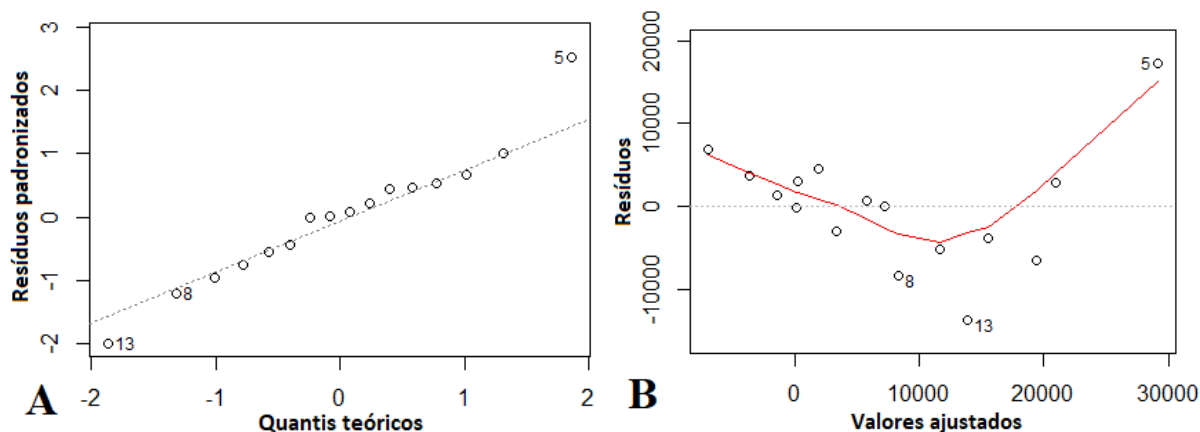
O inverno exibiu as medidas de posição mais baixas dentre as estações, assim como uma assimetria à direita decorrente das observações de *Drosophila* em julho (923) e agosto (2134), na tabela 1, elevarem o valor da média ($\bar{X}=896,50$) em relação à mediana ($Md=102,00$). Por outro lado, a estação de inverno, apresentou a maior dispersão de dados, quando considerado o coeficiente de variação ($CV=182,42\%$), o qual independe do valor da média (Tabela 8). Desta forma, o teste de comparação de grupos pode averiguar a existência de diferenças significativas entre as estações de verão e inverno.

Antes de aplicar ANOVA de blocos aleatorizados para a comparação das médias dos grupos, verificaram-se as suposições de normalidade e homocedasticidade.

O teste Shapiro-Wilk aplicado aos resíduos gerou valores de $W=0,961$ e $p=0,691$, averiguando-se a existência de distribuição normal ($p>0,05$). Este resultado foi confirmado

pelo *qqplot* da figura 6A, na qual os quantis das observações e quantis teóricos apresentaram uma relação linear, havendo poucos desvios da reta.

Figura 6- A. Gráfico *qqplot*; B. Gráfico entre resíduos em relação aos valores ajustados.



Fonte: dados da análise

O teste de Levene aplicado aos resíduos, considerando 3 graus de liberdade, apresentou valor de $F^*=0,477$ e probabilidade do teste maior que $\alpha=0,05$ ($p=0,704$), culminando na aceitação em H_0 sobre a igualdade das variâncias entre as estações. Por outro lado, o gráfico da figura 6B mostrou certa tendência na distribuição de pontos, podendo constituir-se de alguns desvios do pressuposto de homocedasticidade.

Como os testes anteriores revelaram o atendimento dos pressupostos de normalidade e homocedasticidade da ANOVA, ajustou-se o modelo de ANOVA de blocos aleatorizados aos dados, considerando as estações como grupos, e os gêneros, os blocos.

Os resultados do teste ANOVA de blocos aleatorizados aceitaram a hipótese nula sobre a igualdade entre as estações ($F=1,896$ e $p=0,201$), porém, rejeitaram H_0 para os gêneros de moscas ($F=3,984$ e $p=0,046$). Desta forma aplicou o teste de comparações múltiplas de Tukey para determinar quais gêneros de drosofilídeos diferenciavam significativamente entre si.

O teste de Tukey revelou que apenas os gêneros *Zaprionus* e *Rhinoleucophenga* apresentaram diferenças significativas de 20801,25 entre médias ($p=0,0437$). Este resultado foi devido a elevada quantidade de coletas de drosofilídeos do gênero *Zaprionus*, composta apenas pela espécie invasora *Zaprionus indianus*, no decorrer do ano (total de 83339), principalmente em épocas mais quentes, em detrimento das poucas observações de *Rhinoleucophenga*, constituída de 5 espécies, durante o mesmo período (total de 134 coletas).

Estas diferenças no número de coletas dos dois gêneros mencionados podem refletir o nicho ecológico e a capacidade de adaptação as condições adversas das espécies que os compõem. De acordo com Chaves e Tidon (2008), *Z. indianus* constitui-se de uma das espécies de drosofilídeos dominantes do cerrado brasileiro, principalmente na estação chuvosa, decorrente de seus hábitos generalistas, utilizando diversos frutos como substrato, somada a sua alta capacidade de adaptar-se a condições ambientais adversas (MATAVELLI, 2014). Segundo Matavelli (2014), *Z. indianus* possivelmente poderia abrigar em ilhas de calor existentes no meio urbano durante a estação seca, viabilizando sua sobrevivência e a elevação de sua abundância na estação chuvosa.

Diferentemente, o gênero *Rhinoleucophenga* é composto por espécies neotropicais, cujas quantidades de coletas foram bastante reduzidas nos estudos de Coutinho-Silva *et al.* (2017), Hochmüller *et al.* (2010) e Roque, Mata e Tidon (2013), semelhantemente a este trabalho (POPPE; SCHMITZ; VALENTE, 2015). Segundo Roque (2009), diferentemente de drosofilídeos invasores, algumas das espécies neotropicais, principalmente aquelas consideradas raras, são conservadoras quanto ao nicho ecológico, exigentes quanto às condições ambientais e aos recursos para a sobrevivência e reprodução, assim como geralmente apresentam limitações para adaptarem-se as mudanças nos ecossistemas. Desta forma, segundo Mata, Mcgeoch e Tidon (2008), comumente espécies de drosofilídeos nativos são mais abundantes em locais conservados e florestas densas, as quais possibilitam a coexistência de diversos nichos ecológicos, devido às condições diferenciadas de temperatura, umidade e intensidade luminosa em cada um dos seus estratos. Por outro lado, essas espécies seriam geralmente pouco abundantes em espaços alterados pela ação antrópica, como no caso deste trabalho, onde usualmente se predominam condições ambientais mais instáveis (MATA; MCGEOCH; TIDON, 2008).

3.1.3 Correlação entre os gêneros de moscas-das-frutas

Antes da aplicação da análise de correlação, verificou-se a normalidade dos gêneros para direcionar o tipo de correlação mais apropriado. O teste Shapiro-Wilk aplicado aos gêneros rejeitou H_0 acerca da normalidade destes ($p < 0,05$). Considerando que os gêneros não apresentam distribuição normal, optou-se pela utilização do coeficiente de correlação de Spearman.

Tabela 9- Matriz de correlação de Spearman entre as quantidades de coletas por gêneros (*Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*), considerando todos os 12 meses.

Gêneros				
	<i>Zaprionus</i>	<i>Drosophila</i>	<i>Scaptodrosophila</i>	<i>Rhinoleucophenga</i>
<i>Zaprionus</i>	1,000	0,790**	0,643*	0,110
<i>Drosophila</i>	0,790**	1,000	0,482	0,018
<i>Scaptodrosophila</i>	0,643*	0,482	1,000	-0,082
<i>Rhinoleucophenga</i>	0,110	0,018	-0,082	1,000

* $p < 0,05$

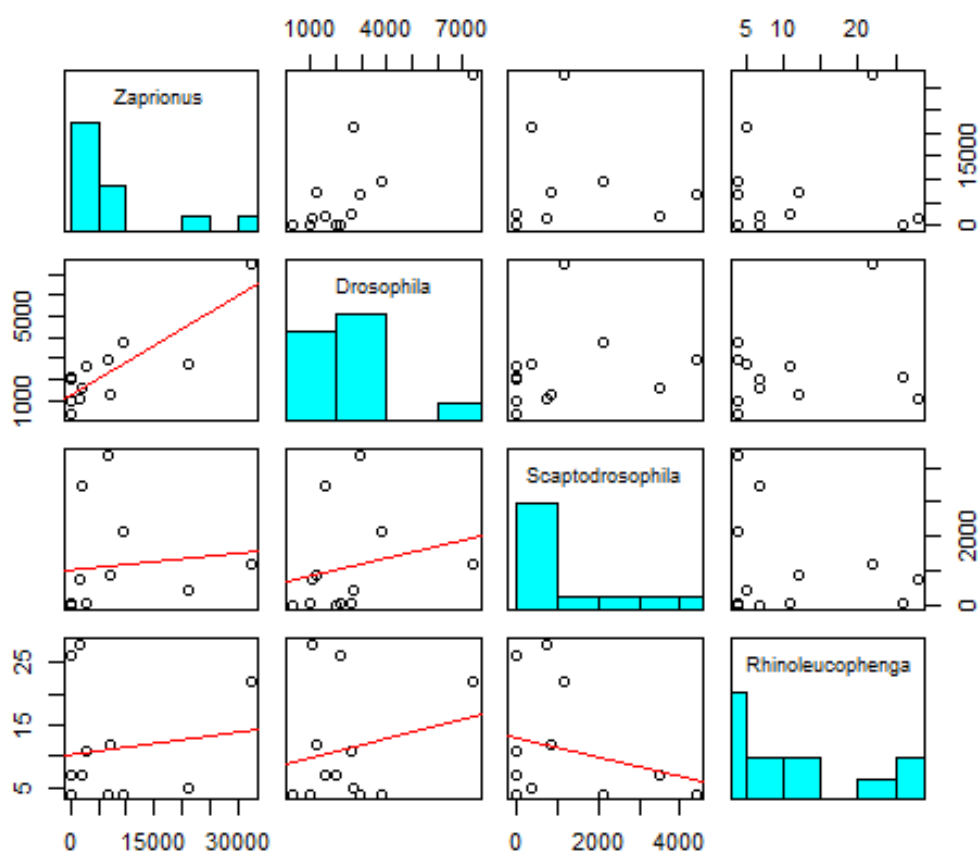
** $p < 0,01$

Fonte: dados da análise

Inicialmente, a tabela 9 informa que se verificaram apenas correlações significativas (coeficiente de correlação para a população $\rho \neq 0$) entre gêneros *Zaprionus* e *Drosophila* ($\rho=0,790$ e $p=0,004$), e *Zaprionus* e *Scaptodrosophila* ($\rho=0,643$ e $p=0,028$), sendo que ambas são positivas.

A figura 7 constitui-se de uma matriz de gráficos de dispersão, na qual cada unidade representa a distribuição dos dados de um gênero (eixo X) em relação a outro (eixo Y), que permite averiguar a presença de associação entre estes, ao observar a existência de orientação dos pontos ou a direção da reta de tendência dos valores. Além disso, a diagonal principal, a qual apresenta a combinação do gênero com ele mesmo gerando o valor do coeficiente de correlação igual a 1, exibe histogramas com as frequências absolutas das coletas por gênero. Nesta matriz, as combinações abaixo e acima da diagonal principal são as mesmas, havendo apenas trocas de eixos dos gêneros.

Figura 7- Matriz de gráficos de dispersão das quantidades de coletas por gêneros (*Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*), considerando todos os 12 meses.



Fonte: dados da análise

Desta forma, observou-se pela matriz gráfica que principalmente os gêneros *Zaprionus* e *Drosophila*, possuíram uma associação positiva, verificada pela direção crescente da reta de tendência dos dados. Essas correlações positivas podem ser explicadas pela tendência dos drosofilídeos se reproduzirem e desenvolverem mais rapidamente durante períodos mais quentes do ano (BONIN, 2010). Os histogramas na diagonal da matriz de gráficos (Figura 7) mostraram que, em geral, os gêneros apresentaram frequências altas de observações baixas ou medianas, e frequências baixas de número de coletas elevados, sendo que estes últimos foram mais restritos às estações mais quentes e úmidas, como verificado nas seções 3.1.1 e 3.1.2.

A inexistência de correlações negativas significativas indicou a possível ausência de competição interespecífica entre os gêneros de drosofilídeos pelos figos, local de oviposição e desenvolvimento dos imaturos, provavelmente pela existência de recursos e nichos ecológicos suficientes no ambiente para permitir a coexistência estável das diferentes espécies de moscas-das-frutas, mesmo que existissem espécies invasoras (principalmente do gênero

Zaprionus) mais aptas a sobreviver e reproduzir nas condições locais (OLIVEIRA, 2010; ROHDE *et al.*, 2010).

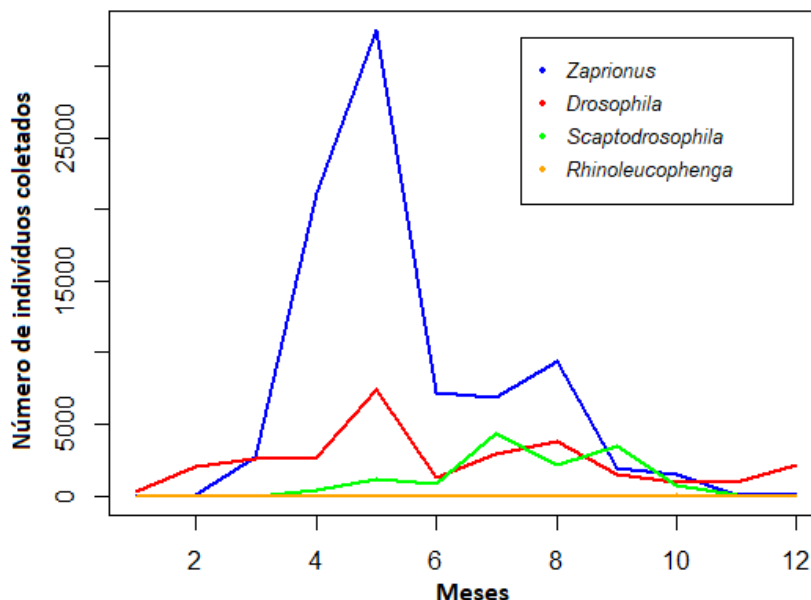
Além disso, Oliveira (2010) ressalta que a coexistência estável de espécies de drosofilídeos poderia também ocorrer em situações, nas quais a competição intraespecífica supera a interespecífica. O autor comenta que este tipo de cenário seria possível por meio da fragmentação no uso do substrato efêmero pelas diferentes espécies de drosofilídeos, de maneira que indivíduos da mesma espécie depositariam seus ovos de forma agregada em partes deste recurso, atraídas por alguma característica dos fragmentos (condições ambientais, estado ou qualidade do substrato), ocasionando uma elevada competição intraespecífica, e reduzida competição interespecífica pelas diversas espécies não compartilharem as mesmas parcelas de recurso (OLIVEIRA, 2010).

De acordo com Oliveira (2010), esta configuração de uso do substrato possibilitaria que as espécies se recuperassem de efeitos negativos da competição intraespecífica, assim como permitiria a coexistência de diferentes espécies, maior exploração do ambiente e colonização de novos nichos, culminando em uma elevação da riqueza local. Porém, a autora comenta sobre a dificuldade de estimar a coexistência por agregação, visto que ainda é um assunto pouco explorado, não se conhecendo completamente os efeitos da sazonalidade e de mudanças ambientais causadas por ações antrópicas neste tipo de interação (OLIVEIRA, 2010).

3.1.4 Análise gráfica de frequências das coletas de moscas-das-frutas

A partir dos valores das frequências absolutas, e do cálculo das frequências relativas, obtidas pela razão do número de coletas mensais de cada gênero em relação ao total de cada mês, construíram-se gráficos de linhas representando a flutuação da abundância no decorrer do ano dos 4 gêneros de drosofilídeos (Figuras 8 e 9).

Figura 8- Gráfico de linhas referentes às frequências absolutas (contagens de indivíduos) dos gêneros *Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*. Meses 1,2...11, e 12 representam respectivamente setembro, outubro,...julho, e agosto.



Fonte: dados da análise

Observando-se o gráfico da figura 8, verificou-se que os gêneros *Zaprionus* e *Drosophila* possuíram as maiores frequências absolutas no mês de janeiro (mês 5), apresentando um pequeno pico no mês de abril (mês 8). Os valores de coletas maiores no mês de janeiro poderiam ser consequência das temperaturas e umidade mais elevadas que favorecem a produção de figos, assim como a reprodução e o desenvolvimento do drofílídeos (BONIN, 2010; CARRER; GARCIA, 2007; NOVAIS; BRITO, 2017).

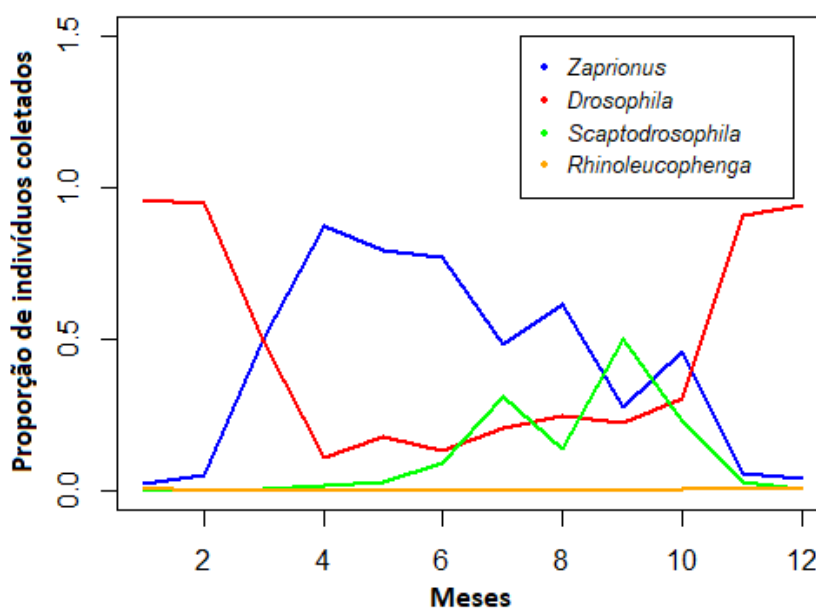
Semelhantemente aos gêneros *Zaprionus* e *Drosophila*, as maiores frequências de *Scaptodrosophila* ocorreram em março e maio (meses 7 e 9), que correspondem ao final do verão (estação chuvosa) e início do outono (estação seca) (CARRER; GARCIA, 2007; NOVAIS; BRITO, 2017). Apesar dos meses de abril e maio constituírem do início do outono e da estação de seca do clima subtropical de inverno seco e verão úmido, a ocorrência do segundo pico para os três gêneros considerados pode ser decorrente da elevada produção de sincônios pelas figueiras, a qual geralmente se estende pelos meses de novembro à maio (ALVARES *et al.*, 2014; MATAVELLI, 2014; ZAMPAULO *et al.*, 2007).

Para estes três gêneros com maiores valores de coletas, verificou-se também uma redução abrupta do número de observações em fevereiro (mês 6), em relação aos meses anteriores (Figura 8). Esta diminuição pode ser consequência de mudanças ambientais atípicas

não favoráveis para o desenvolvimento dos drosofilídeos, como quedas na temperatura e períodos de estiagem (BONIN, 2010; ROHDE *et al.*, 2010).

Quanto ao gráfico da figura 9 percebeu-se que o gênero *Zaprionus* apresentou frequências relativas elevadas durante os meses de dezembro até abril (meses 4 e 8, respectivamente), devido aos altos valores de coletas (6840 até 32476), passando a reduzir consideravelmente (abaixo de 100 indivíduos) em julho e agosto (meses 11 e 12), semelhantemente a *Scaptodrosophila*. Por outro lado, durante os períodos de setembro e outubro (meses 1 e 2), e julho e agosto (meses 11 e 12), as proporções relativas de *Drosophila* aumentaram em relação as demais, devido a menor variação do número de coletas que *Zaprionus* e *Scaptodrosophila*, como pode ser verificado no gráfico da figura 8.

Figura 9- Gráfico de linhas referentes às frequências relativas (proporção de coletas por mês) dos gêneros *Zaprionus*, *Drosophila*, *Scaptodrosophila* e *Rhinoleucophenga*. Meses 1,2...11, e 12 representam respectivamente setembro, outubro,...julho, e agosto.



Fonte: dados da análise

O gênero *Rhinoleucophenga* como apresentou baixo número de coletas nos meses (variou entre 4 e 28 observações) comparados aos três maiores gêneros, teve suas frequências absolutas e relativas mascaradas em ambos gráficos da figuras 8 e 9. Segundo os dados da tabela 6, as maiores abundâncias de espécimes de *Rhinoleucophenga* ocorreram nas estações de outono (39 indivíduos) e verão (38 indivíduos), ou na estação seca (73 indivíduos coletados de abril à setembro), quando considerado o clima subtropical úmido de inverno seco e verão quente, sendo que a maior observação ocorreu no mês de junho, como constatado na

tabela 1 (ALVARES *et al.*, 2014; CARRER; GARCIA, 2007; NOVAIS; BRITO, 2017; ZAMPAULO *et al.*, 2007).

O padrão de abundância encontrado neste trabalho para o gênero *Zaprionus*, que neste estudo é representado por apenas uma espécie, *Z. indianus*, é semelhante aquele verificado pelo estudo Chaves e Tidon (2008), onde a espécie em questão, juntamente com *Drosophila simulans* (Sturtevant), constituam-se as espécies mais abundantes do cerrado brasileiro. Outras pesquisas apresentaram padrões semelhantes ao encontrados nestas análises, como nos estudos de Tidon, Leite e Leão (2003) e Hochmüller *et al.* (2010), em que *Z. indianus* predominava durante a estação chuvosa, sendo mais abundante em vegetações abertas ou áreas de cultivo, quando comparado com fragmentos de florestais. Diferentemente, o trabalho de Castro e Valente (2001) averiguou que esta espécie apenas ocorria durante os meses de dezembro à maio em Porto Alegre (RS), englobando estações de verão e outono, possuindo maior abundância na área urbana.

Neste estudo, o gênero *Drosophila* constitui-se do mais rico, contando com 51 espécies, tanto neotropicais quanto exóticas. Dentre as espécies exóticas, destacou-se *D. simulans*, a qual compunha 72,28% das coletas de moscas-das-frutas para este grupo, e 16,52% do total de indivíduos coletados. Este resultado também foi observado no trabalho de Roque, Mata e Tidon (2013), em que o gênero *Drosophila* apresentou 51 espécies coletadas em áreas de florestas de galeria de Brasília (DF), sendo que a espécie *D. simulans* constituiu de 28% do total de coletas, as quais foram mais abundantes na estação seca, diferentemente do resultado desta análise.

Alguns estudos, semelhantemente a esta análise, verificaram a presença de *D. simulans* no decorrer de todo o ano, tanto em áreas de cultivo quanto em regiões de mata ripária, como o trabalho de Hochmüller *et al.* (2010), enquanto em outros, esta espécie foi mais abundante em estações chuvosas e em áreas menos degradadas, como na pesquisa de Castro e Valente (2001). Por outro lado, o estudo de Torres e Ravazzi (2006) verificou que *D. simulans* apresentava elevação de sua abundância durante as transições de temperaturas entre as estações, tanto em espaços urbanos quanto em fragmentos de floresta semidecidual do noroeste do estado de São Paulo.

Diferentemente deste trabalho, o estudo de Coutinho-Silva *et al.* (2017), em fragmentos de mata atlântica da região pernambucana, verificou que as espécies *Z. indianus* e *Scaptodrosophila latifasciaeformis* (Duda) foram mais abundantes na estação seca, enquanto *D. simulans* apresentou maior abundância relativa durante a estação chuvosa. A espécie *S. latifasciaeformis*, a única representante do gênero neste trabalho, teve poucos indivíduos

coletados no decorrer do ano, restringindo as coletas aos meses de fevereiro à abril (final da estação chuvosa e início da seca) na pesquisa de Roque, Mata e Tidon (2013), a qual exibiu um padrão de flutuação de abundância pouco semelhante a encontrada nesta análise.

No estudo de Roque, Mata e Tidon (2013) também se verificou a ocorrência de espécimes do gênero *Rhinoleucophenga* no dossel das matas de galeria do Centro-Oeste brasileiro, o qual apresentou poucas coletas no decorrer do período, compondo menos de 1% do total, não ocorrendo preferência por estações. Semelhantemente, o trabalho de Hochmüller *et al.* (2010), a única espécie deste gênero, *Rhinoleucophenga obesa* (Loew), registrou apenas uma coleta na área conservada. Neste estudo, semelhantemente ao de Coutinho-Silva *et al.* (2017), o gênero *Rhinoleucophenga* exibiu a segunda maior quantidade de espécies (3 spp), depois de *Drosophila* (29 spp), porém possuiu poucos espécimes coletados nos fragmentos, apresentando maior abundância geral na estação seca (13 indivíduos).

Considerando as semelhanças e diferenças de padrões de flutuação das populações de drosofilídeos desta análise com os estudos apresentados anteriormente, Torres e Ravazzi (2006) comentam da dificuldade de atribuir a variação de sua abundância no decorrer do ano somente a fatores abióticos (temperatura e precipitação) ou a fatores bióticos, tornando a análise mais complexa.

Complementando a discussão sobre a complexidade da flutuação de drosofilídeos no decorrer do ano, Matavelli (2014) e Roque, Mata e Tidon (2013) comentam que as dessemelhanças de padrões entre os estudos poderiam ser devido às diferenças de condições ambientais (temperatura, precipitação, sazonalidade) dos locais pesquisados, proporcionados pelos distintos climas e tipos vegetacionais que compõem os biomas (Mata Atlântica, Cerrado, entre outros), somada as particularidades da composição faunística de cada um destes espaços. As diferenças nas comunidades florísticas e faunísticas poderiam proporcionar que as espécies de drosofilídeos ocupassem nichos ecológicos distintos em diferentes ambientes, ao procurarem por condições de sobrevivência mais favoráveis, evitando competição com outras espécies por abrigo e substrato, ou adaptando-se as condições locais, característica comum em espécies generalistas e invasoras (MATAVELLI, 2014; ROQUE; MATA; TIDON, 2013)

Desta forma, as particularidades das condições ambientais e das comunidades de seres vivos de cada local, dificultam a comparação dos resultados de estudos realizados em diferentes tipos vegetacionais ou com diferentes níveis de conservação. Sendo assim, releva-se a importância do investimento em pesquisas que busquem avaliar a flutuação de drosofilídeos em áreas agrícolas de culturas diversas e em diferentes regiões do país, visto

que, seus padrões podem diferenciar daqueles encontrados em áreas conservadas, fragmentos e espaços urbanos, possibilitando uma melhor compreensão da dinâmica destes insetos em ambientes diversos (COUTINHO-SILVA *et al.*, 2017; ROQUE; MATA; TIDON, 2013).

3.2 Análise dos dados referentes às posturas de *C. megacephala* em diferentes densidades

Nesta segunda seção apresentam-se os resultados das análises referentes à comparação entre grupos (potes) de mesma réplica, os quais estão organizados e divididos segundo a densidade de califorídeos adultos de *C. megacephala* correspondente (100, 200, 400 e 800 dípteros por gaiola ou réplica). Ao final da seção referente à densidade 800 califorídeos por réplica, discutem-se os resultados principais de forma conjunta.

3.2.1 Posturas de *C. megacephala* na densidade 100 califorídeos por réplica

A densidade 100 califorídeos por gaiola apresentou duas réplicas, cujas análises consideraram as contagens dos dias 11 à 66, para a primeira réplica, e dias 10 à 66, para a segunda, decorrente da exclusão dos primeiros dias com observações iguais a zero.

Em ambas as réplicas observaram-se assimetrias positivas ou à direita ($g_1 > 0$), com médias aritméticas superiores aos valores das medianas (Tabelas 10 e 11). Este fato ocorreu devido à existência de zeros antes do primeiro quartil (25% dos dados) nos potes 2 e 4 na primeira réplica, e pote 4, na segunda, sendo que nos restantes dos grupos a predominância de zeros alcança até a mediana (50% dos valores). A grande quantidade de zeros, somada a presença de poucos valores elevados, como medidas discrepantes representadas por pequenos círculos vazios nos *boxplots* das figuras 10A e 10B, causaram as elevações das médias em relação às medianas. A maior parte dos potes das réplicas da densidade 100 califorídeos por gaiola apresentou curtose do tipo leptocúrtica ($g_2 > 3$), onde existe elevada densidade de probabilidades nas caudas da distribuição em relação a sua região central (GOTELLI; ELLISON, 2011).

Tabela 10- Análise exploratória das posturas nos potes da primeira réplica da densidade 100 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	56	56	56	56
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	0,00	34,00	0,00	271,50
Terceiro Quartil (75%)	227,25	306,75	267,25	676,00
Máximo	1144,00	2350,00	1395,00	1564,00
Média Aritmética	124,63	258,80	203,52	409,89
Variância	61060,53	193564,60	121108,18	174741,99
Desvio Padrão	247,10	439,96	348,01	418,02
Coefficiente de Variação	198,28%	170,00%	170,99%	101,98%
Assimetria (g₁)	2,37	2,77	1,94	0,91
Curtose (g₂)	8,56	11,90	5,95	3,15

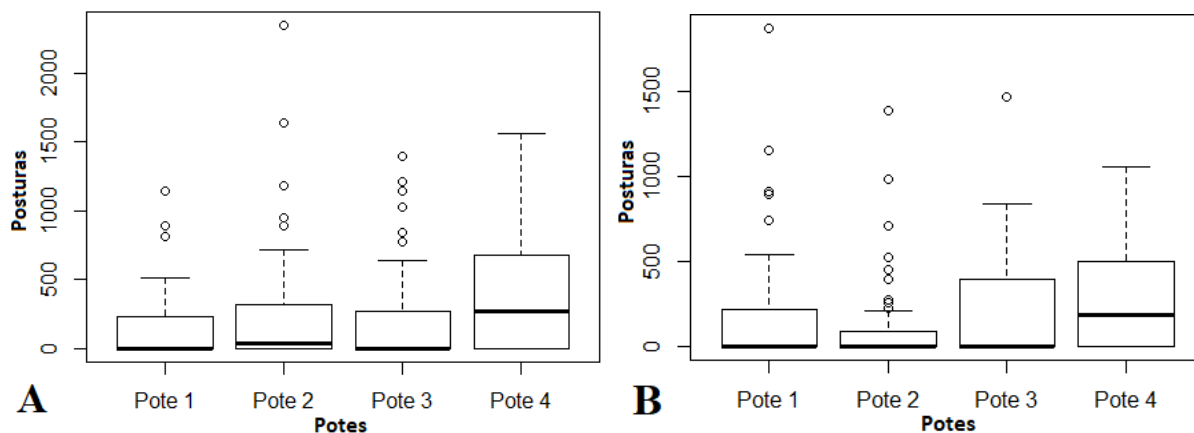
Fonte: dados da análise

Tabela 11- Análise exploratória das posturas nos potes da segunda réplica da densidade 100 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	57	57	57	57
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	0,00	0,00	0,00	184,00
Terceiro Quartil (75%)	220,00	90,00	400,00	504,00
Máximo	1875,00	1386,00	1471,00	1058,00
Média Aritmética	168,21	111,16	215,53	288,21
Variância	123258,10	66758,56	95498,25	118554,38
Desvio Padrão	351,08	258,38	309,03	344,32
Coefficiente de Variação	208,72%	232,44%	143,38%	119,47%
Assimetria	2,87	3,19	1,68	0,93
Curtose	12,22	13,92	6,16	2,52

Fonte: dados da análise

Figura 10- *Boxplots* das posturas nos potes das réplicas de densidade 100 califorídeos por gaiola.
A. Primeira réplica; **B.** Segunda réplica.



Fonte: dados da análise

Os potes 4 da primeira e segunda réplicas, apresentaram distribuições de dados próximos da mesocúrtica ($g_2 \cong 3$) e platicúrtica ($g_2 < 3$), respectivamente (Tabelas 10 e 11). No caso da curva mesocúrtica, esta apresenta distribuição de dados semelhante à curva normal (gaussiana), enquanto a segunda apresenta reduzidas densidades de probabilidades dos valores extremos em relação aos medianos (AYRES, 2011; GOTELLI; ELLISON, 2011).

Em ambas as réplicas verificaram-se que os potes 4 apresentaram maiores valores de mediana ($Md=271,50$ e $Md=184,00$, para a primeira e segunda réplica respectivamente) e média ($\bar{X}=409,89$ e $\bar{X}=288,21$). Observou-se também que os potes, os quais exibiram menores valores para as médias (potes 1 e 2, respectivamente, para a primeira e segunda réplica), possuíam maiores valores de coeficientes de variação ($CV=198,28\%$ e $CV=232,44\%$), diferentemente das variâncias ($s^2=61060,53$ e $s^2=66758,56$) e dos desvios padrões ($s=247,10$ e $s=258,38$), os quais foram as menores dentre os grupos considerados (Tabelas 10 e 11).

Considerando as análises obtidas pela descrição breve dos dados, o teste de comparação de grupos permitiu averiguar a existência de diferenças significativas entre o pote 4 com os restantes dos grupos, para cada réplica de forma isolada.

O teste de Kruskal-Wallis rejeitou a hipótese nula de igualdade dos potes tanto para primeira quanto a segunda réplica do experimento para a densidade 100 califorídeos por gaiola ($p=0,0005$ e $p=0,013$, respectivamente), gerando valores de $H=17,904$ e $H=10,778$, considerando 3 graus de liberdade.

Nas comparações múltiplas de Student-Newman-Keuls verificou que a rejeição da hipótese H_0 no teste anterior ocorreu pela existência de diferenças significativas entre o pote 4 com os demais, para a primeira réplica, e entre o pote 4 com grupos 1 e 2, para a segunda

réplica. Através dos testes de comparações múltiplas averiguou-se que em ambos os casos, os potes 4, além de apresentarem maiores médias e medianas, exibiam também maiores valores das somas de ranks (7891,00 e 7677,00, para primeira e segunda réplica, respectivamente).

3.2.2 Posturas de *C. megacephala* na densidade 200 califorídeos por réplica

A densidade 200 califorídeos por gaiola possui duas réplicas, nas quais foram consideradas na análise o número de posturas dos dias 9 à 85, para a primeira réplica, e dos dias 11 à 80, para a segunda.

Em ambas as réplicas, os dados aproximaram de uma curva do tipo leptocúrtica ($g_2 > 3$), e exibiram assimetria positiva, decorrente de pelo menos 25% das medidas se constituírem de zeros, para os potes 3 e 4, se estendendo para 50% nos demais (Tabelas 12 e 13). A presença de observações discrepantes contribuiu também para elevação dos valores das médias em relação as suas medianas (Figuras 11A e 11B).

Tabela 12- Análise exploratória das posturas nos potes da primeira réplica da densidade 200 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	77	77	77	77
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	0,00	0,00	232,00	90,00
Terceiro Quartil (75%)	247,00	285,00	575,00	461,00
Máximo	1453,00	1615,00	2188,00	2685,00
Média Aritmética	197,39	197,95	384,58	315,21
Variância	116550,87	97393,21	263358,80	246630,85
Desvio Padrão	341,40	312,08	513,18	496,62
Coefficiente de Variação	172,96%	157,66%	133,44%	157,55%
Assimetria	2,03	2,08	1,84	2,38
Curtose	6,57	7,95	6,48	9,68

Fonte: dados da análise

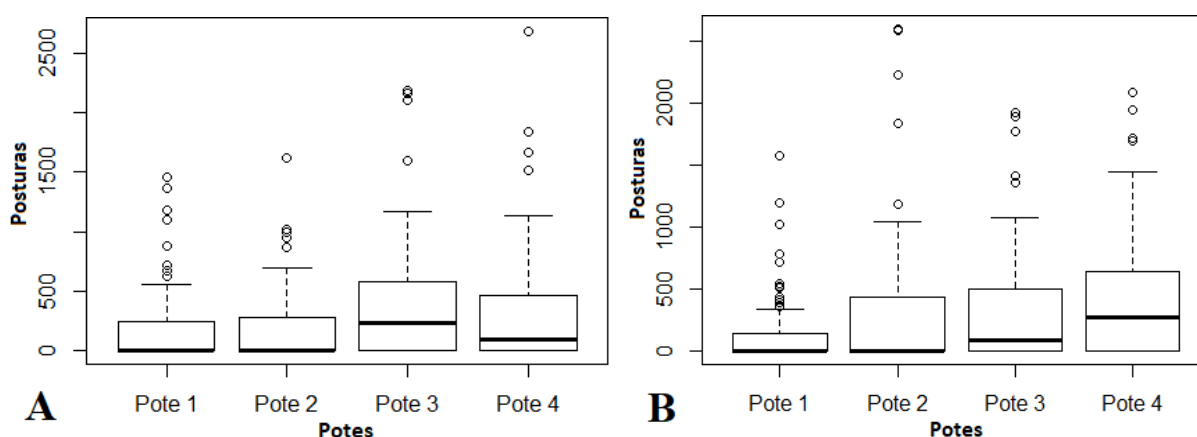
Tabela 13- Análise exploratória das posturas nos potes da segunda réplica da densidade 200 califórdeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	70	70	70	70
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	0,00	0,00	85,00	273,00
Terceiro Quartil (75%)	136,25	435,75	497,75	628,00
Máximo	1580,00	2602,00	1928,00	2092,00
Média Aritmética	147,129	331,314	325,457	404,429
Variância	93892,78	341419,70	229278,45	244364,71
Desvio Padrão	306,42	584,31	478,83	494,33
Coefficiente de Variação	208,27%	176,36%	147,13%	122,23%
Assimetria	2,69	2,49	1,82	1,62
Curtose	10,67	9,18	5,85	5,43

Fonte: dados da análise

Figura 11- Boxplots das posturas nos potes das réplicas de densidade 200 califórdeos por gaiola.

A. Primeira réplica; **B.** Segunda réplica.



Fonte: dados da análise

Na primeira réplica (Tabela 12), observou-se que o pote 3 apresentou valores mais elevados de mediana ($Md=232,00$) e média ($\bar{X}=384,58$). Na réplica considerada, o primeiro pote exibiu a maior dispersão dos dados comparados aos demais ($CV=172,96\%$), porém com os segundos menores valores para variância ($s^2=116550,87$) e desvio padrão ($s=341,40$), por considerarem a reduzida média do grupo ($\bar{X}=197,39$).

Na segunda réplica (Tabela 13), verificou-se que o pote 4 exibiu maiores valores para mediana ($Md=273,00$) e média ($\bar{X}=404,429$). Semelhantemente a primeira réplica, o primeiro pote possuiu maior valor para o coeficiente de variação ($CV=208,27\%$), porém apresentou menores valores para variância ($s^2=93892,78$) e desvio padrão ($s=306,42$).

Considerando-se a análise preliminar dos dados da densidade 200 califorídeos por gaiola, esperou-se encontrar, com o teste de comparação de grupos, diferenças significativas entre o pote 3 com os demais, para a primeira réplica, e entre o grupo 4 com os restantes, para a segunda.

O teste de Kruskal-Wallis apenas rejeitou a hipótese nula de igualdade entre os potes para a segunda réplica da densidade 200 ($p=0,0049$), gerando valores de $H=12,861$ para 3 de graus de liberdade.

A rejeição de H_0 para a segunda réplica ocorreu devido a existência de diferenças significativas entre o pote 1 com os grupos 3 e 4, como observado nos testes de comparações múltiplas de Student-Newman-Keuls. Assim como verificado na análise exploratória, o pote 1 da réplica considerada, além de apresentar menores média ($\bar{X}=147,128$) e mediana ($Md=0$), exibiu menor soma de ranks (8048,50), a qual foi significativamente diferente dos grupos 3 e 4, os quais, por outro lado, possuíram as maiores médias ($\bar{X}=325,457$ e $\bar{X}=404,428$, respectivamente), medianas ($Md=85,00$ e $Md=273,00$) e somas de ranks (10221,00 e 11424,50).

3.2.3 Posturas de *C. megacephala* na densidade 400 califorídeos por réplica

A densidade 400 califorídeos por gaiola apresentou quatro réplicas, nas quais foram consideradas para a análise as posturas dos dias 10 à 80, 9 à 85, 8 à 75, e 11 à 66, para as primeira, segunda, terceira e quarta réplicas, respectivamente.

Em geral, observou-se que as réplicas apresentaram dados com assimetrias positivas ($g_1 > 0$), e cujas curtoses aproximaram-se de curvas do tipo leptocúrtica ($g_2 > 3$). Além disso, verificou-se também que a maior parte das réplicas possuiu potes, os quais exibiram os primeiros quartis ou inferiores (25%) compostos por zeros. Alguns grupos estenderam este número até a mediana (50%), como os potes 2 da primeira réplica, 1 e 2 da terceira e todos os potes da quarta, sendo que este último, possuiu seu grupo 2 com 75% dos dados (terceiro quartil ou superior) compostos por zeros (Tabelas 14, 15, 16 e 17). Também se averiguou a presença de inúmeras observações discrepantes através dos gráficos das figuras 12A, 12B, 12C e 12D, as quais foram responsáveis por elevar os valores das médias de cada pote em relação às suas medianas.

Tabela 14- Análise exploratória das posturas nos potes da primeira réplica da densidade 400 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	71	71	71	71
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	208,00	0,00	190,00	242,00
Terceiro Quartil (75%)	686,00	429,00	595,00	705,50
Máximo	2150,00	2567,00	1910,00	1970,00
Média Aritmética	402,93	287,93	378,93	394,94
Variância	286318,24	225363,69	212655,95	207255,77
Desvio Padrão	535,09	474,72	461,15	455,25
Coefficiente de Variação	132,80%	164,88%	121,70%	115,27%
Assimetria	1,48	2,51	1,20	1,23
Curtose	4,53	10,62	3,78	4,15

Fonte: dados da análise

Tabela 15- Análise exploratória das posturas nos potes da segunda réplica da densidade 400 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	77	77	77	77
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	200,00	216,00	199,00	217,00
Terceiro Quartil (75%)	471,00	447,00	636,00	765,00
Máximo	2416,00	1501,00	2323,00	2290,00
Média Aritmética	358,32	284,00	436,00	476,58
Variância	285442,12	135980,16	329946,63	310759,17
Desvio Padrão	534,27	368,75	574,41	557,46
Coefficiente de Variação	149,10%	129,84%	131,75%	116,97%
Assimetria	2,19	1,57	1,51	1,30
Curtose	7,91	5,16	4,42	4,18

Fonte: dados da análise

Tabela 16- Análise exploratória das posturas nos potes da terceira réplica da densidade 400 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	68	68	68	68
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	0,00	0,00	274,50	179,00
Terceiro Quartil (75%)	462,75	422,75	841,75	671,00
Máximo	4134,00	1884,00	3863,00	3170,00
Média Aritmética	357,46	304,44	623,66	432,06
Variância	482290,61	237652,94	773749,87	371225,61
Desvio Padrão	694,47	487,50	879,63	609,28
Coefficiente de Variação	194,28%	160,13%	141,04%	141,02%
Assimetria	3,09	1,84	1,84	1,88
Curtose	14,91	5,41	5,96	7,46

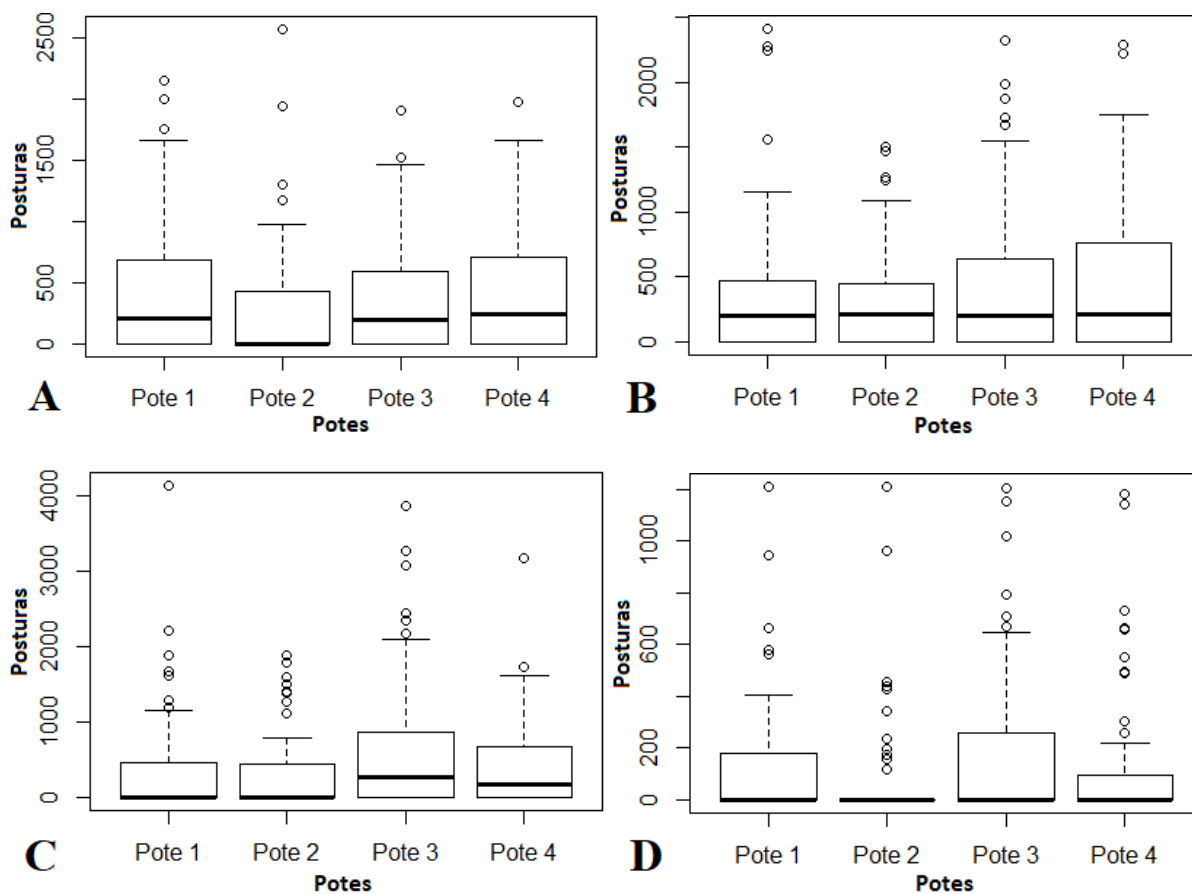
Fonte: dados da análise

Tabela 17- Análise exploratória das posturas nos potes da quarta réplica da densidade 400 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	56	56	56	56
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	0,00	0,00	0,00	0,00
Terceiro Quartil (75%)	173,50	0,00	241,75	47,25
Máximo	1211,00	1208,00	1205,00	1183,00
Média Aritmética	126,48	94,91	191,46	138,07
Variância	61940,25	53672,77	91024,36	82594,18
Desvio Padrão	248,88	231,67	301,70	287,39
Coefficiente de Variação	196,77%	244,10%	157,58%	208,15%
Assimetria	2,55	3,20	1,92	2,19
Curtose	9,72	13,87	6,09	7,12

Fonte: dados da análise

Figura 12- *Boxplots* das posturas nos potes das réplicas de densidade 400 califórídeos por gaiola.
A. Primeira réplica; **B.** Segunda réplica; **C.** Terceira réplica; **D.** Quarta réplica.



Fonte: dados da análise

Na primeira réplica (Tabela 14) verificou-se que o pote 1 apresentou maior média ($\bar{X}=402,93$), porém não a maior mediana ($Md=208,00$), a qual pertenceu ao pote 4 ($Md=242,00$). O primeiro pote também exibiu dados bastante dispersos em relação aos demais, possuindo o segundo maior valor do coeficiente de variação ($CV=132,80\%$) e as maiores variância ($s^2=286318,24$) e desvio padrão ($s=535,09$).

Na segunda réplica (Tabela 15), percebeu-se que o pote 4 possuiu maiores valores para média ($\bar{X}=476,58$) e mediana ($Md=217,00$), porém a menor dispersão dentre os grupos, quando considerado o coeficiente de variação ($CV=116,97\%$), diferentemente de quando se considera a variância ($s^2=310759,17$) e desvio padrão ($s=557,46$) em torno da média, possuindo as segundas maiores medidas.

Na terceira réplica (Tabela 16), averiguou-se maiores valores de média ($\bar{X}=623,66$) e mediana ($Md=274,50$) para o pote 3, o qual apresentou a maior dispersão dos dados quando considerada a variância ($s^2=773749,87$) e o desvio padrão ($s=879,63$) em torno da média,

porém exibiu o segundo menor coeficiente de variação entre os grupos considerados ($CV=141,04\%$).

Na quarta réplica (Tabela 17), observou-se que o pote 3 também apresentou os valores mais elevados de média ($\bar{X}=191,46$), variância ($s^2=91024,36$) e desvio padrão ($s=301,70$), porém a menor dispersão de dados quando considerado o coeficiente de variação ($CV=157,58\%$).

Considerando os grupos que exibiram maiores valores de média e mediana em cada réplica, esperou-se encontrar diferenças significativas com os potes com menores medidas destes parâmetros.

O teste de Kruskal-Wallis rejeitou H_0 sobre a igualdade dos dados somente para a terceira réplica da densidade 400 califorídeos por gaiola ($p=0,0416$), na qual se obteve o valor de $H= 8,226$ para 3 de graus de liberdade.

A rejeição da hipótese nula para a terceira réplica foi decorrente da presença de diferenças significativas entre o pote 3 com os grupos 1 e 2, como verificado no teste de comparações múltiplas de Student-Newman-Keuls. O pote 3, além de apresentar maiores valores de média e mediana, como visto na descrição inicial, exibiu também a mais elevada soma de ranks (10660,50) dentre os grupos da terceira réplica.

3.2.4 Posturas de *C. megacephala* na densidade 800 califorídeos por réplica

A densidade 800 califorídeos por gaiola possuiu três réplicas, em que, para a análise, foram consideradas as posturas dos dias 8 à 74, na primeira réplica, e 8 à 77, para a segunda e terceira.

As três réplicas, em geral, são caracterizadas por exibirem assimetria positiva dos dados ($g_1 > 0$), e uma curtose do tipo leptocúrtica ($g_2 > 3$). Nas tabelas 18, 19 e 20, averiguou-se também que as medidas inferiores ao primeiro quartil (25% dos dados) foram compostas por zeros, exceto para os potes 2 da primeira e terceira réplica, na qual este valor se estendeu até a mediana (50% dos dados). As inúmeras observações iguais a zero, somada a existência de dados discrepantes, como verificado nas figuras 13A, 13B e 13C, contribuíram para que as médias assumissem valores superiores as medianas.

Tabela 18- Análise exploratória das posturas nos potes da primeira réplica da densidade 800 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	67	67	67	67
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	84,00	0,00	147,00	74,00
Terceiro Quartil (75%)	514,50	268,00	527,00	361,00
Máximo	1812,00	1778,00	2471,00	1764,00
Média Aritmética	311,85	188,96	338,42	267,30
Variância	204201,37	108806,65	272848,85	176504,36
Desvio Padrão	451,89	329,86	522,35	420,12
Coefficiente de Variação	144,90%	174,57%	154,35%	157,17%
Assimetria	1,59	2,48	2,48	1,98
Curtose	4,71	10,32	9,76	6,43

Fonte: dados da análise

Tabela 19- Análise exploratória das posturas nos potes da segunda réplica da densidade 800 califorídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	69	69	69	69
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	115,00	28,00	215,00	180,00
Terceiro Quartil (75%)	365,00	239,00	445,00	355,00
Máximo	1975,00	1764,00	2082,00	2013,00
Média Aritmética	254,93	170,07	360,97	293,71
Variância	140001,98	75481,80	211388,65	148902,33
Desvio Padrão	374,17	274,74	459,77	385,88
Coefficiente de Variação	146,77%	161,54%	127,37%	131,38%
Assimetria	2,41	3,31	1,67	2,15
Curtose	10,22	18,07	5,50	8,31

Fonte: dados da análise

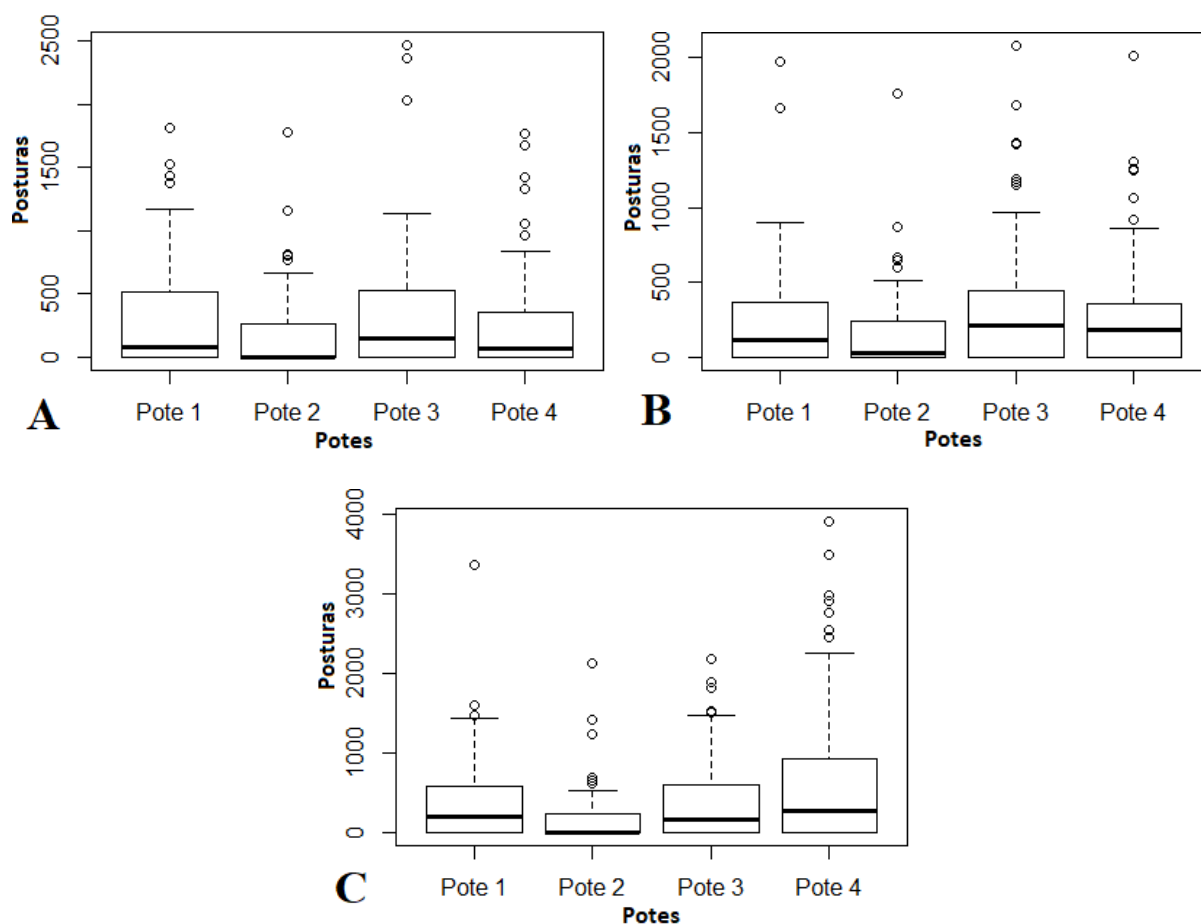
Tabela 20- Análise exploratória das posturas nos potes da terceira réplica da densidade 800 califórídeos por gaiola.

Parâmetros	Pote 1	Pote 2	Pote 3	Pote 4
Tamanho amostral	70	70	70	70
Mínimo	0,00	0,00	0,00	0,00
Primeiro Quartil (25%)	0,00	0,00	0,00	0,00
Mediana	201,50	0,00	166,50	265,00
Terceiro Quartil (75%)	562,50	224,50	587,50	912,25
Máximo	3355,00	2115,00	2185,00	3905,00
Média Aritmética	408,33	174,96	393,87	717,03
Variância	328600,72	129198,91	301520,84	929901,56
Desvio Padrão	573,24	359,44	549,11	964,31
Coefficiente de Variação	140,39%	205,45%	139,41%	134,49%
Assimetria	2,46	3,33	1,57	1,59
Curtose	11,44	15,88	4,54	4,63

Fonte: dados da análise

Figura 13- Boxplots das posturas nos potes das réplicas de densidade 800 califórídeos por gaiola.

A. Primeira réplica; **B.** Segunda réplica; **C.** Terceira réplica.



Fonte: dados da análise

Observou-se na primeira réplica (Tabela 18) que o pote 3 exibiu os valores mais elevados de mediana ($Md=147,00$), média ($\bar{X}=338,42$), variância ($s^2=272848,85$) e desvio padrão ($s=522,35$), porém, se considerar o valor do coeficiente de variação ($CV=154,35\%$), este apresentaria a segunda menor dispersão de dados da réplica em questão.

Na segunda réplica (Tabela 19), o pote 3 também apresentou os maiores valores de mediana ($Md=215,00$), média ($\bar{X}=360,97$), variância ($s^2=211388,65$) e desvio padrão ($s=459,77$), porém a menor dispersão de dados entre os grupos, quando considerado o coeficiente de variação ($CV=127,37\%$).

Na terceira réplica (Tabela 20), o pote 4 exibiu os valores mais elevados para mediana ($Md=265,00$), média ($\bar{X}=717,03$), variância ($s^2=929901,56$) e desvio padrão ($s=964,31$), por outro lado, possuiu a menor dispersão quando considerado o coeficiente de variação ($CV=134,49\%$), o qual não depende do valor da média.

Considerando a análise preliminar dos dados da densidade 800 califorídeos por gaiola, esperou-se encontrar diferenças significativas entre os potes com maiores e menores valores de mediana e média, no caso, o pote 3 com os restantes, na primeira e segunda réplicas, e o grupo 4 com os demais, na terceira réplica.

O teste de Kruskal-Wallis rejeitou a hipótese nula sobre igualdade dos grupos apenas para a terceira réplica ($p=0,0002$), obtendo-se valor de $H=19,60$ para 3 graus de liberdade. A rejeição de H_0 para a terceira réplica ocorreu devido a existência de diferenças significativas entre o pote 2 com os demais grupos, como observado no teste de comparações múltiplas de Student-Newman-Keuls. Na réplica considerada, o pote 2 consistiu no grupo que apresentou menores valores de mediana ($Md=0$) e média ($\bar{X}=174,96$), que influenciou na sua reduzida soma de postos (7454,00), a qual diferenciou significativamente das somas de ranks dos potes 1, 3 e 4 (10300,50, 9994,50 e 11591,00, respectivamente).

A existência de diferenças significativas observadas em algumas das comparações anteriores, principalmente na primeira e segunda réplica da densidade 100, e na terceira réplica da densidade 400 califorídeos por gaiola, nas quais algumas posições apresentavam número de posturas significativamente maior que outras, pode ser explicada pela tendência de *C. megacephala* realizar postura agregada, na qual fêmeas desta espécie são atraídas por feromônios liberados durante a oviposição de outras, e acabam por efetuar a postura no mesmo substrato (NORRIS 1965; VON ZUBEN; STANGENHAUS; GODOY, 2000). Segundo Rivers, Thompson e Brogan (2011), a postura agregada favoreceria a formação de massa de imaturos, a qual conferiria vantagens à sobrevivência das larvas. Dentre estas vantagens, pode-se destacar: geração de temperatura interna na massa de imaturos

favorecendo a aceleração do metabolismo e o desenvolvimento das larvas; facilitação da digestão de nutrientes pela liberação conjunta de enzimas digestivas e ação cooperativa dos ganchos bucais durante o consumo do substrato; proteção contra a dessecação e contra parasitos ou predadores (CORRÊA, 2013; RIVERS; THOMPSON; BROGAN, 2011).

Por outro lado, as análises anteriores verificaram a existência de diferenças significativas somente em 5 réplicas de um total 11 (45,45%), que corresponderam a 100% daquelas da densidade 100 califorídeos por gaiola, 50% em 200, 25% em 400, e 33,33% em 800. Desta forma, verificou-se neste estudo uma tendência mais uniforme da distribuição do número de posturas entre os potes à medida que se aumentou a densidade de moscas nas réplicas, ao invés de uma oviposição agregada como sugerido inicialmente.

Alguns estudos como de Lima e Von Zuben (2016) sugerem que fêmeas de *C. megacephala* poderiam realizar a avaliação prévia do substrato, preferindo efetuar a oviposição separadamente de outras massas de ovos, buscando evitar superfícies saturadas, onde ocorreria competição intraespecífica por alimento. Segundo o trabalho de Von Zuben *et al.* (1993) a competição intraespecífica por alimento devido às altas densidades lavais, reduz a quantidade de alimento disponível para cada imaturo, diminuindo sua sobrevivência, desacelerando o desenvolvimento larval, gerando adultos com dimensões e fecundidade reduzidas, afetando, desta forma, o sucesso reprodutivo dos descendentes de *C. megacephala*. Além da competição intraespecífica, Rivers, Thompson e Brogan (2011) sugerem que a oviposição agregada de *C. megacephala* também poderia ser desvantajoso para as larvas, decorrente da atração de predadores e parasitos pelos feromônios liberados durante a postura, e do estresse causado pelas temperaturas mais elevadas no interior das massas de imaturos, que quando ultrapassado um limite de temperatura favorável, poderia prejudicar o desenvolvimento larval.

Quanto considerado os grupos que apresentaram maiores médias nas comparações, verificou-se que o pote 1 compôs 9,09% do total de 11 réplicas, enquanto os potes 3 e 4 perfizeram 45,45%, cada um. Por outro lado, quando se considera apenas as réplicas que exibiram diferenças significativas entre os grupos, observou-se a predominância de valores mais elevados de médias, medianas e somas de ranks em potes na posição 4, os quais representaram 80% das réplicas, enquanto as posições 1 e 2 possuíram os menores valores para estas medidas em 60% e 40% das réplicas consideradas, respectivamente. Estes resultados demonstraram uma preferência de *C. megacephala* pela oviposição no substrato na posição 4, e uma certa precaução com as posições 1 e 2, diferentemente da hipótese inicialmente sugerida, em que a postura agregada ocorreria de forma aleatória em substratos

supostamente idênticos, apenas variando quanto ao seu posicionamento. Segundo Corrêa (2013) a preferência por uma determinada posição para a postura de ovos, pode estar associada a alguma condição ambiental (temperatura, luminosidade, umidade) ou do substrato (composição, estado) heterogênea desconhecida no experimento, que atraiu as fêmeas de *C. megacephala* para realizarem as oviposições em um determinado pote.

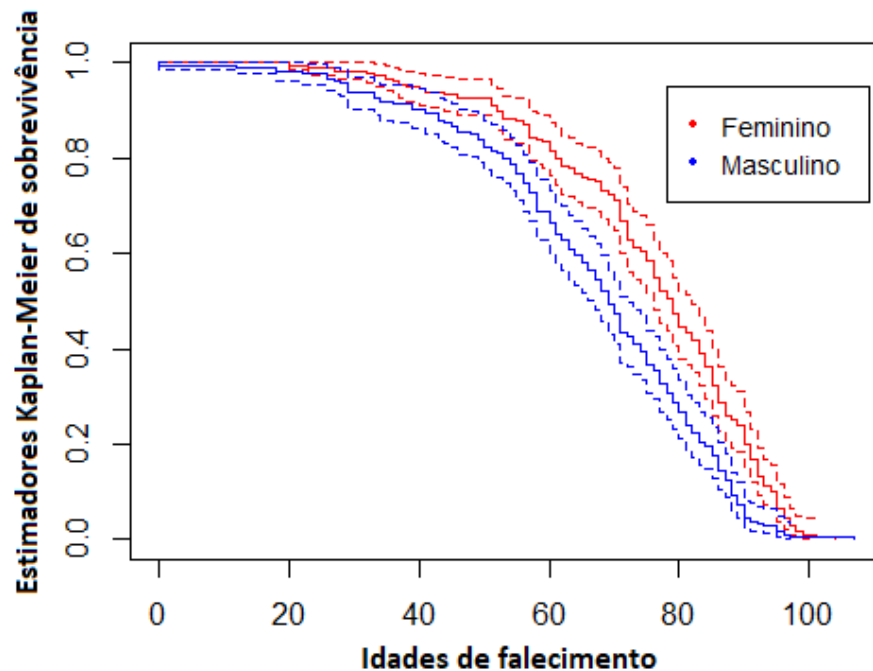
Deve-se considerar também que houve desvios nas tendências, como visto na porcentagem da densidade 800 (33,33%) de diferenças significativas das comparações de grupos, serem superiores daquelas de 400 califorídeos por gaiola (25%). Este desvio pode ser ocasionado pelo reduzido número de réplicas para cada uma das densidades (2 para 100 e 200, 4 para 400, e 3 para 800), visto que a aceitação da igualdade em uma das réplicas de 200 califorídeos por gaiola, reduziu em 50% a ocorrência de diferenças significativas entre os potes. Desta forma, a baixa quantidade de replicação do experimento pode dificultar a realização de inferências conclusivas sobre o fenômeno estudado, porém nem sempre é viável proceder com elevado número de réplicas necessárias para obtenção de maior confiabilidade de resultados, decorrentes de limitações materiais, financeiras e de tempo de estudo (GOTELLI; ELLISON, 2011).

3.3 Análise de sobrevivência dos dados de Bebedouro 2019

A partir dos dados necrológicos do município de Bebedouro, SP, realizou-se o cálculo do estimador de sobrevivência acumulada de Kaplan-Meier para cada sexo e idade de falecimento, obtendo-se as curvas do gráfico da figura 14. No apêndice I são apresentados os valores dos tempos de falhas, número de sobreviventes sob risco, quantidade de falhas, probabilidades e taxas de falhas, estimadores de sobrevivência acumulada e seu intervalo de confiança de 95% para cada uma das idades de falecimento e dos sexos, assim como também são exibidos as razões entre taxas de falhas multiplicadas por 1000 da população masculina em relação à feminina.

Pode-se observar com a figura 14 que a sobrevivência da população masculina do município reduziu mais rapidamente com o avanço da idade, quando comparado a feminina, porém atingiu o valor zero posteriormente, pela maior idade de falha do sexo masculino ser 107 anos, enquanto o feminino foi 104 anos. Por outro lado, a maior parte dos dados referentes aos falecimentos no município em questão referiu-se aos indivíduos do sexo masculino (53,25%) de um valor total de 385 óbitos computados no período de 1 de janeiro à 31 de dezembro de 2019.

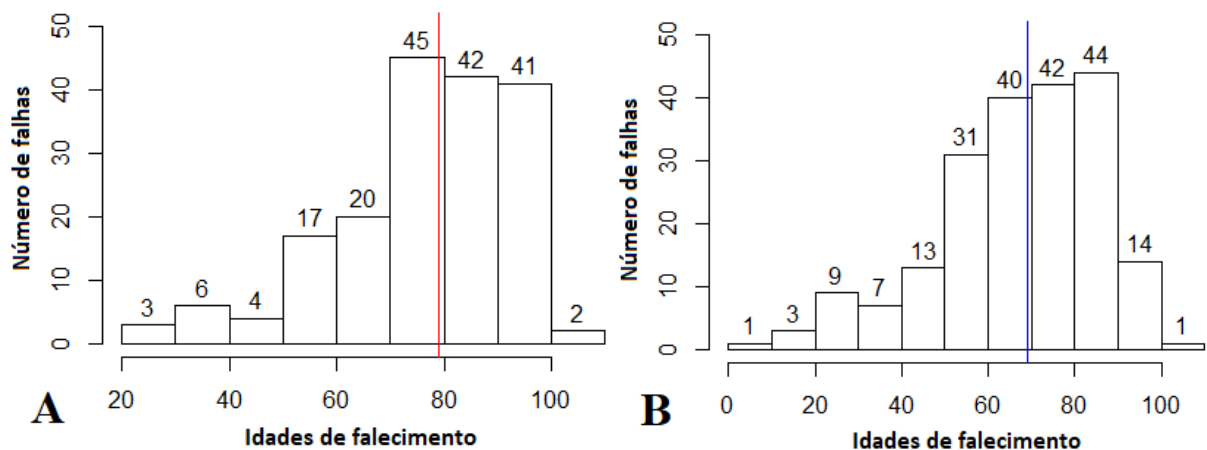
Figura 14- Curvas de sobrevivência estratificadas por sexo em relação às idades de falecimento dos indivíduos, com seus respectivos intervalos de confianças de 95% (linhas tracejadas).



Fonte: dados da análise

Além disso, verificou-se que o tempo mediano de vida (t_{md}) e mediana apresentaram menores valores para o sexo masculino ($t_{md}=68,76$ e $Md=69$ anos), em relação ao feminino ($t_{md}=78,18$ e $Md=79$ anos), sendo que este último parâmetro está destacado por uma linha vertical colorida dos histogramas das figuras 15A e 15B. O tempo mediano de vida constitui-se de um parâmetro que expressa a idade a qual a sobrevivência acumulada atinge o valor de 0,5 ou 50% da inicial.

Figura 15- Histogramas de frequências absolutas de falhas para ambos os sexos, com linha vertical representando a mediana do conjunto. **A.** Sexo feminino; **B.** Sexo masculino.

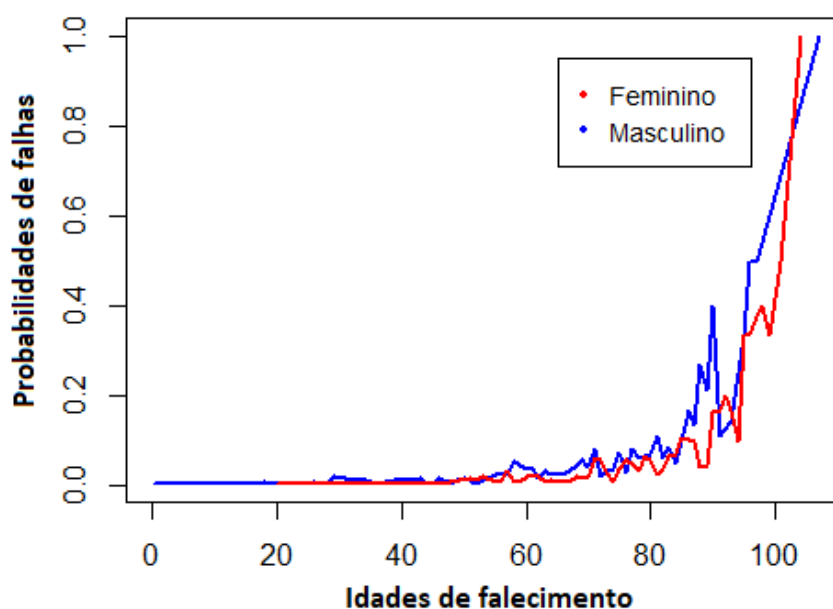


Fonte: dados da análise

Nos histogramas das figuras 15A e 15B, verificou-se a ausência de óbitos para o sexo feminino em idades anteriores a 20 anos, diferentemente do masculino, o qual apresentou 4 falhas. Além disso, observou-se que a população masculina apresentou número de falhas (64) maior que o dobro da feminina (30) para idades anteriores à 60 anos. Desta forma, a quantidade de ocorrências superior para o sexo masculino em idades menos avançadas influenciou na queda mais acentuada de sua curva de sobrevivência e de seu tempo de médio de vida. Por outro lado, também pode se averiguar nos histogramas que o maior número de falhas para o sexo feminino ocorreu entre o intervalo de 70 à 79 anos (45 ocorrências), passando a diminuir posteriormente, enquanto na população masculina, a quantidade mais elevada de ocorrências sucedeu-se nas idades entre 80 à 89 anos (44 falhas).

Quanto às probabilidades de falha em certa idade (Figura 16), verificou-se valores de probabilidade de 0,2 até 0,4 para as idades 92, 95 à 99, atingindo 0,5 e 1,0 em 101 e 104 anos, para a população feminina. No sexo masculino, as probabilidades de falhas superiores a 0,2 ocorreram anteriormente aos 90 anos, como nas idades 88 à 90, e 95, alcançando aos 96, 97 e 107 anos probabilidades de 0,5, 0,5 e 1,0.

Figura 16- Gráfico de probabilidades de falhas para ambos os sexos em relação às idades de falecimento.

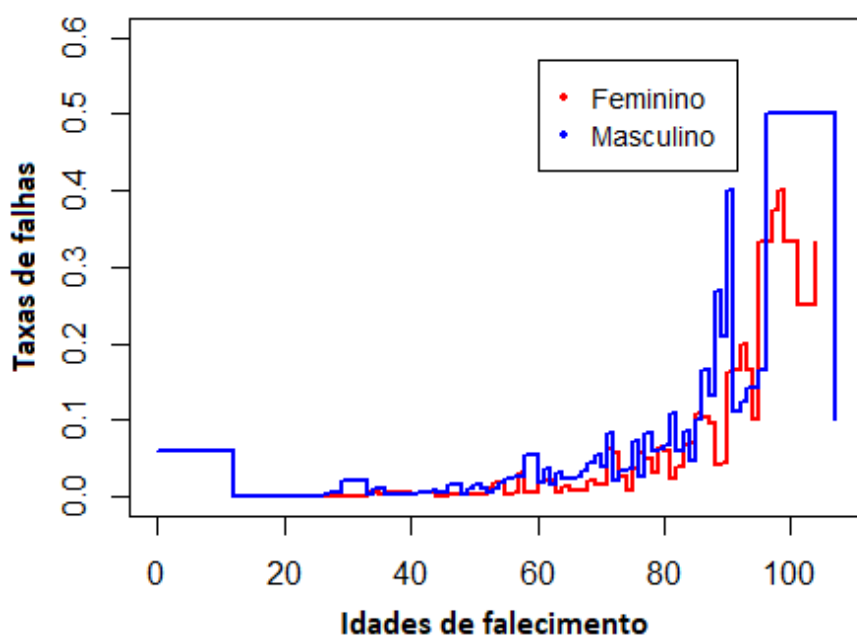


Fonte: dados da análise

No gráfico da figura 17, o qual apresenta as taxas de falhas para o sexo feminino e masculino, observou-se constância nos valores de taxas nos intervalos entre idades de falecimento distintas, por representarem as quantidades de ocorrências por ano, no período

considerado. Desta forma, averiguou-se para população feminina baixas taxas de falhas (0 à 0,015) antes dos 60 anos, exceto para as idades 53 à 54, e 57 (0,019 e 0,032 falhas por ano, respectivamente), decorrente do número de ocorrências entre 3 e 5 nestes períodos. Observou-se também que as maiores taxas de falhas (acima de 0,1 falhas por ano) para o sexo feminino ocorreram a partir dos 85 anos, intercalados por valores mais reduzidos, alcançando taxas superiores a 0,3 nas idades 95 à 99 e 104 anos.

Figura 17- Gráfico de taxas de falhas para ambos os sexos em relação às idades de falecimento.



Fonte: dados da análise

Considerando o gráfico da figura 17, diferentemente da população feminina, a masculina apresentou algumas elevações em idades anteriormente aos 60 anos, como as taxas falhas entre 0,015 à 0,03 nas idades 29 à 33, 46 à 47, 50 e 54 à 57 anos, e taxas superiores à 0,05 nas idades 0,08 (1 mês) à 12, e 58 anos. Averiguou-se também que as maiores taxas de falha (acima de 0,1 falhas ao ano) para o sexo masculino ocorreram a partir dos 81 anos, intercalados por valores menores, alcançando taxas superiores a 0,3 nas idades 90, 96 à 106 anos.

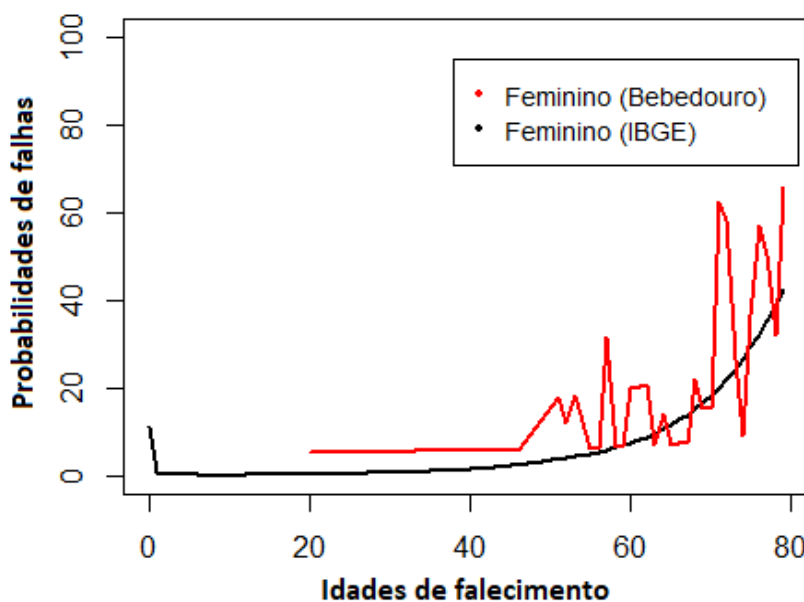
Buscando verificar melhor a ocorrência de sobremortalidade masculina, realizou-se a razão das taxas de falhas para as duas populações considerando todas as idades de falecimento (apêndice I), averiguando-se diferenças mais evidentes em idades abaixo dos 60 anos. As taxas de falhas foram de 3,5 à 5 vezes maiores para o sexo masculino em relação o feminino nas idades de 27 à 28, 32, 34, 43, 49, 50, 55 e 57 anos, passando a serem 5,5 à 8,2 vezes mais

elevados em 19 à 21, 46 à 47, 54 e 58 anos, e atingindo valores entre 14 e 18 vezes em idades de 29 à 30 anos.

Os valores de probabilidades e taxas de falhas diferenciaram daqueles encontrados nos histogramas, por considerarem o número de sobreviventes até o momento, aumentando os seus valores em idades mais avançadas decorrente da reduzida quantidade de indivíduos vivos no momento, ocasionada por complicações naturais relacionadas ao envelhecimento (ABREU; CÉSAR; FRANÇA, 2009; IBGE, 2019).

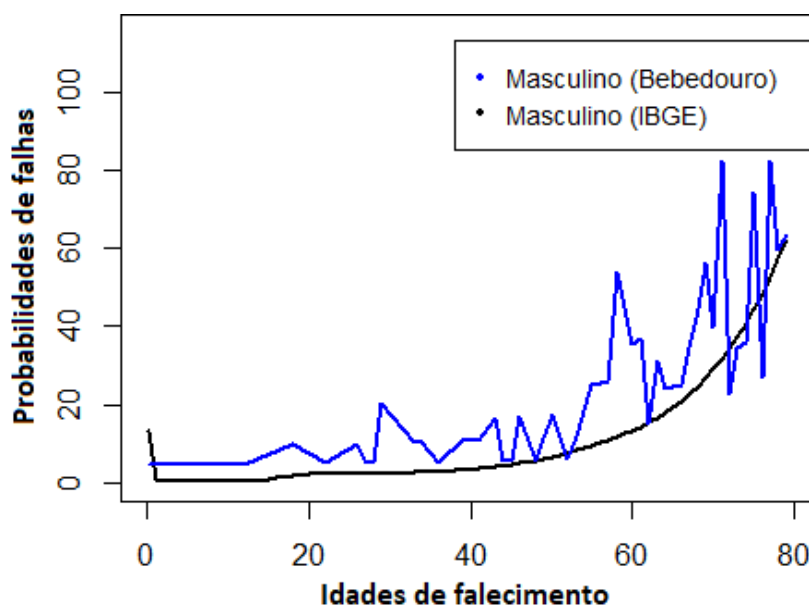
As figuras 18 e 19 contrapõem as probabilidades de falhas a cada 1000 habitantes, para populações femininas e masculinas brasileiras, referentes aos dados do IBGE de 2018, com as probabilidades de ocorrência do município de Bebedouro multiplicado por 1000, a qual permitiu melhor comparação dos dados. Nesta figuras foram consideradas apenas as probabilidades de falhas até a idade 79 anos de ambos conjuntos de dados, visto que a tábua de mortalidade do IBGE considera o valor de probabilidade igual a 1000 para idades iguais ou maiores que 80 anos.

Figura 18- Gráfico de probabilidades de falhas para o sexo feminino por 1000 habitantes para dados de Bebedouro e IBGE, considerando idades de falha até 79 anos.



Fonte: dados da análise

Figura 19- Gráfico de probabilidades de falhas para o sexo masculino por 1000 habitantes para dados de Bebedouro e IBGE, considerando idades de falha até 79 anos.



Fonte: dados da análise

A partir destes gráficos (Figuras 18 e 19), percebeu-se que, em geral, as probabilidades de falhas do Brasil aumentaram progressivamente com o avanço da idade (exceção primeiros anos de vida e acima dos 80 anos), diferentemente das probabilidades de falhas de Bebedouro, as quais apresentaram grande variação nos seus valores, principalmente nas idades mais avançadas (a partir dos 50 anos).

Considerando os intervalos de 20 à 58 anos para a população feminina, e 12 à 46 anos para a masculina, averiguou-se que as probabilidades de falhas de Bebedouro superaram os valores referentes ao Brasil. Após os intervalos de idades mencionados até 79 anos, verificou-se intercalação das maiores probabilidades destes conjuntos de dados, devido a grande variação destes valores para o município de Bebedouro. Esta grande variação ocorreu em decorrência do conjunto de dados analisados não se constituir de uma grande amostra, comparada a toda população brasileira, sendo que pequenas diferenças no número de ocorrências de Bebedouro alteraram intensamente os valores de probabilidades de falhas calculadas.

Como verificado tanto nos dados do IBGE quanto nos dados de Bebedouro, as mais altas probabilidades e taxas de falhas ocorreram em idades mais avançadas, decorrente ao grande avanço da medicina e da indústria farmacêutica desde a década de 1940, que iniciou o processo de transição demográfica, marcada primeiramente pela redução da mortalidade, principalmente no início da vida (0-5 anos). Progressivamente, com a melhoria da renda,

saneamento básico, sistema de saúde (atendimentos, campanhas de vacinação, acompanhamento pré-natal) e educação, intensificou-se a redução da mortalidade e natalidade, ocorrendo também à elevação da expectativa de vida. Neste contexto, também houve a transição epidemiológica, em que as causas de mortes por infecção e parasitoses, foram substituídas por problemas associados ao envelhecimento, como cardiopatias e câncer (IBGE, 2019).

Considerando os primeiros anos de vida, a tábua de mortalidade do IBGE apresentou valores mais elevados de probabilidades de falhas principalmente na idade 0 (11,351 a cada 1000 habitantes para sexo feminino, e 13,305 para o masculino), se estendendo até os 5 ou 6 anos, reduzindo posteriormente, a partir do qual os valores passaram a aumentar progressivamente, como verificado no restante dos gráficos (Figura 18 e 19). Este padrão foi verificado também na população masculina de Bebedouro, porém a probabilidade de falha apresentou valor inferior (4,9 por 1000 habitantes), a qual representava apenas uma única ocorrência neste período na idade 1 mês ou 0,08 ano. Diferentemente, o sexo feminino não apresentou nenhuma falha neste intervalo, se mantendo com probabilidades nulas até a idade de 20 anos.

O relatório do IBGE (2019) verificou que a sobremortalidade masculina concentrou nas idades entre 15 aos 29 anos (probabilidades de falhas entre 1,024 e 2,337 por 1000 habitantes), na população adultos jovens. Enquanto a população masculina de Bebedouro, as maiores probabilidades de falhas da idade adulta ocorreram em idades mais avançadas, entre 26 à 58 anos, com probabilidades entre 10 à 54 ocorrências por 1000 habitantes. Quando considerada as razões das taxas de falhas entre as duas populações de Bebedouro, verificaram-se maiores diferenças entre taxas em idades mais avançadas, principalmente nos 19 aos 28 anos, com as taxas de ocorrências da população masculina aproximadamente 3,5 à 6 vezes maior que a feminina, atingindo as mais elevadas diferenças (superior a 14 vezes) entre 29 aos 31 anos.

Buscando verificar-se a diferença da sobrevivência observada na figura 14 era significativa, aplicou-se o teste Log-rank. O seguinte teste gerou um valor de qui-quadrado de 23,9 e probabilidade de teste menor que $\alpha=0,05$ ($p=1 \cdot 10^{-6}$), rejeitando a hipótese nula sobre a igualdade da sobrevivência entre os sexos.

De maneira geral, o resultado deste conjunto de análises (curva de sobrevivência, histogramas, probabilidades, taxas de falhas, e teste de hipótese) mostrou que a população de Bebedouro-SP seguiu a tendência comum verificada na população brasileira, quando considerados os sexos, na qual geralmente observa-se mortalidade superior na população

masculina. Este resultado comumente observado é ocasionado por fatores externos ou não naturais (homicídios, acidentes de trânsito, entre outros) presentes em sociedades urbanas, os quais atingem principalmente jovens adultos do sexo masculino (ABREU; CÉSAR; FRANÇA, 2009; IBGE, 2019). Além disso, Abreu, César e França (2009) acrescentam que diferenças na mortalidade entre os sexos podem ser decorrentes das características socioeconômicas das populações, ofertas de serviços de saúde locais, questões biológicas e socioculturais sobre a prevenção, diagnóstico e tratamento precoce, sendo que estes últimos fatores estão comumente mais presentes na população feminina.

4 CONCLUSÃO

Conclui-se, a partir deste trabalho, que o conhecimento e as ferramentas estatísticas fornecem importante contribuição para a compreensão dos processos biológicos. Os testes de comparação de grupos e os gráficos de frequências absolutas e relativas, aplicados ao primeiro conjunto de dados, permitiram verificar a existência de diferenças significativas das coletas de drosofilídeos entre os meses, provavelmente devido ao maior número de coletas em períodos mais quentes e úmidos, os quais favoreceriam a frutificação do figo e desenvolvimento destes insetos. Por outro lado, observou-se diferenças nos padrões de flutuação de abundância das moscas nos gráficos de frequências absolutas e relativas do estudo em relação à literatura, decorrente das particularidades das condições ambientais e das comunidades de seres vivos das regiões pesquisadas, as quais possivelmente influenciam na composição e nos nichos ecológicos dos drosofilídeos locais. Neste primeiro conjunto, a análise de correlação averiguou a existência de poucas associações positivas significativas entre os gêneros, sugerindo coexistência estável entre estes, ao invés de competição interespecífica como presumido inicialmente.

No segundo conjunto, as comparações de grupos revelaram a ocorrência de diferenças significativas em menos da metade das réplicas, as quais corroboram com a informação comum da literatura acerca da postura agregada de *C. megacephala*. Em contrapartida, a tendência de elevação da igualdade entre grupos à medida que se aumentava a densidade de califorídeos da réplica, apóia a teoria de avaliação do substrato pela fêmea, a qual evita, desta maneira, a competição intraespecífica dos imaturos, que se constitui em um dos fatores que interferem no sucesso reprodutivo de seus descendentes.

No terceiro conjunto de dados, as curvas de sobrevivência, probabilidades e taxas de falhas, histogramas e teste de hipóteses mostraram a existência de diferenças significativas

entre a sobrevivência da população feminina e masculina de Bebedouros, SP, fato que poderia estar associado à sobremortalidade masculina decorrente de causas não naturais (acidentes, violências, entre outros). Além disso, observou-se também a existência de algumas semelhanças entre as tendências crescentes das curvas de probabilidade de falha com o aumento da idade, do município estudado com os dados referente à população brasileira em 2018.

Desta forma, neste estudo, verificou-se que as ferramentas estatísticas forneceram resultados que muitas vezes corroboram com teorias consolidadas na biologia, porém também revelou alguns padrões diferentes daqueles comumente ressaltados pela literatura, os quais podem abrir caminhos para a exploração destas possibilidades que vão muito além das análises e conclusões feitas neste trabalho, como a questão da coexistência estável, a flutuação de drosofilídeos em cultivos diversos, e a hipótese de avaliação do substrato por *C. megacephala*.

5 REFERÊNCIAS

ABREU, D. M. X.; CÉSAR, C. C.; FRANÇA, E. B. Diferenciais entre homens e mulheres na mortalidade evitável no Brasil (1983-2005). **Cad. Saúde Pública**, Rio de Janeiro, v. 25, n. 12, p. 2672-2682, 2009. Disponível em: <https://www.scielo.org/pdf/csp/2009.v25n12/2672-2682/pt>. Acesso em: 19 mar. 2020.

ALVARES, C. A. *et al.* Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, [s.l.], v. 22, n. 6, p. 711-728, 2014. Disponível em: http://www.lerf.eco.br/img/publicacoes/Alvares_etal_2014.pdf. Acesso em: 21 mar. 2020.

AYRES, M. **Elementos da Bioestatística: A Seiva do Açaizeiro**. Belém: Universidade Federal do Pará, 2011. 551 p.

AYRES, M *et al.* **BIOESTAT – Aplicações estatísticas nas áreas das ciências bio-médicas**. Versão 5.3. Belém: Ong Mamiraua. 2007. Disponível em: <http://www2.assis.unesp.br/ffrei/bioestat.html>. Acesso em: 20 fev. 2020.

BACH, R. **Manual do Messias: Lembretes para a alma avançada**. Tradução: Julia Hellmuth. Rio de Janeiro: Record, 2006. 221 p.

BAUMGARTNER, D. L.; GRENNBERG, B. The genus *Chrysomya* (Diptera: Calliphoridae) in the New World. **Journal of Medical Entomology**, [s.l.], v. 21, n. 1, p.105-113, 1984.

Disponível em: <https://academic.oup.com/jme/article-abstract/21/1/105/2220221?redirectedFrom=fulltext>. Acesso em: 17 jun. 2020.

BOAVENTURA, D. D. **Desenvolvimento de *Zaprionus indianus* Gupta, 1970 (Diptera: Drosophilidae) em diferentes dietas**. 2013. 54 f. Trabalho de Conclusão de Curso (Bacharelado em Ciências Biológicas) - Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, 2013.

BONIN, T. G. **Variação na curva de crescimento e tamanho de asa de *Zaprionus indianus* (Gupta, 1970) (Diptera: Drosophilidae) de acordo com diferentes temperaturas**. 2010. 49 f. Trabalho de Conclusão de Curso (Bacharelado em Ecologia) - Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, 2010.

CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artmed, 2003, 255 p.

CARRER, T. T.; GARCIA, A. Classificação climática para a cidade de Ituverava, SP. **Revista Nucleus**, [s.l.], v. 4, n. 1-2, p. 29-38, 2007. Disponível em: <http://www.nucleus.feituverava.com.br/index.php/nucleus/article/view/9/36>. Acesso em: 19 mar. 2019.

CARVALHO, M. S.; CAMPOS, D. P.; OLIVEIRA, R. V.C. **Introdução a análise de sobrevivência**. [200-?]. Disponível em: <http://sobrevida.fiocruz.br/material/slidesintro.pdf>. Acesso em: 20 jan. 2020.

CASTRO, F. L.; VALENTE, V. L. S. *Zaprionus indianus* is invading Drosophilid communities in the southern Brazilian city of Porto Alegre. **Dros. Inf. Serv.**, [s.l.], v. 84, [s.n.], p. 15-17, 2001. Disponível em: <http://www.ou.edu/journals/dis/DIS84/R7%20Castro/Castro.pdf>. Acesso em: 8 abr. 2020.

CHAVES, N. B.; TIDON, R. Biogeographical aspects of drosophilids (Diptera, Drosophilidae) of the Brazilian savanna. **Rev. Bras. entomol.**, São Paulo, v. 52, n. 3, p. 340-348, 2008. Disponível em: <http://www.scielo.br/pdf/rbent/v52n3/a05v52n3.pdf>. Acesso em: 8 abr. 2020.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. São Paulo: Edgard Blücher, 2006, 367 p.

CORRÊA, L. F. **Aspectos semioquímicos e visuais no comportamento de ovipostura em *Chrysomya megacephala* (Fabricius) (Diptera: Calliphoridae)**. 2013. Dissertação (Mestrado em Zoologia)- Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, 2013. Disponível em:

<https://repositorio.unesp.br/bitstream/handle/11449/108744/000766949.pdf?sequence=1&isAllowed=y>. Acesso em: 04 abr. 2020.

COUTINHO-SILVA, R.D. *et al.* Effects of seasonality on drosophilids (Insecta, Diptera) in the northern part of the Atlantic Forest, Brazil. **Bulletin of Entomological Research**, [s.l.], v. 107, n. 5, p. 1- 11, 2017. Disponível em: <https://www.cambridge.org/core/journals/bulletin-of-entomological-research/article/effects-of-seasonality-on-drosophilids-insecta-diptera-in-the-northern-part-of-the-atlantic-forest-brazil/C863D601539C708EC4D3C7073FA0DE46>. Acesso em: 2 abr. 2019.

FOX, J.; WEISBERG, S. An {R} Companion to Applied Regression. Third Edition. Thousand Oaks CA: Sage. 2019. Disponível em: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>

GOMES, L.; VON ZUBEN, C. J.; SANCHES, M. R. Estudo da dispersão larval radial pós-alimentar em *Chrysomya megacephala* (Fabricius) (Diptera, Calliphoridae). **Rev. Bras. entomol.**, São Paulo, v. 47, n. 2, p. 229-234, 2003. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0085-56262003000200012. Acesso em: 15 ago. 2019.

GOTELLI, N. J.; ELLISON, A. M. **Princípios de estatística em ecologia**. Tradução: Fabrício Beggato Baccaro. Porto Alegre: Artmed, 2011, 528 p.

HOCHMÜLLER, C. J. C. *et al.* The drosophilid fauna (Diptera, Drosophilidae) of the transition between the Pampa and Atlantic Forest Biomes in the state of Rio Grande do Sul, southern Brazil: first records. **Papéis Avulsos Zoologia**, São Paulo, v. 50, n. 19, p. 285-295, 2010. Disponível em: <http://www.scielo.br/pdf/paz/v50n19/a01v5019.pdf>. Acesso em: 15 ago. 2019.

IBGE. **Tábua completa de mortalidade para o Brasil – 2018:**

Breve análise da evolução da mortalidade no Brasil. Rio de Janeiro, 2019. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/periodicos/3097/tcmb_2018.pdf. Acesso em: 19 mar. 2020.

KOMSTA, L.; NOVOMESTKY, F. moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14: [s.l.]. 2015. Disponível em: <https://CRAN.R-project.org/package=moments>.

LEONEL, S.; SAMPAIO, A. C. (Orgs.). **A figueira**. São Paulo: Unesp, 2011. 395 p.

LIMA, T.; VON ZUBEN, C. J. *Chrysomya megacephala* (Fabricius) (Diptera: Calliphoridae) Oviposition Behavior in Previous Oviposition Situation. **Neotrop Entomol**, [s.l.], v. 45, [s.n.], p. 612–617, 2016. Disponível em: <https://link.springer.com/article/10.1007/s13744-016-0412-z>. Acesso em: 04 abr. 2020.

MANLY, B. J. F. **Métodos Estatísticos Multivariados: Uma Introdução**. 3 ed. Tradução: Sara Ianda Carmona. Porto Alegre: Bookman, 2008, 229 p.

MATA, R. A.; MCGEOCH, M.; TIDON, R. Drosophilid assemblages as a bioindicator system of human disturbance in the Brazilian Savanna. **Biodivers Conserv**, [s.l.], v.17, [s.n.], p.2899–2916, 2008. Disponível em: <https://link.springer.com/article/10.1007/s10531-008-9403-7>. Acesso em: 13 abr. 2020.

MATAVELLI, C. **Dinâmica populacional de *Zaprionus indianus* Gupta 1970 (Diptera: Drosophilidae) e caracterização de alguns aspectos biológicos**. 2014. Dissertação (Doutorado em Zoologia)- Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, 2014. Disponível em: <https://repositorio.unesp.br/handle/11449/121886>. Acesso em: 09 abr. 2020.

MICROSOFT. **Microsoft Office Excel 2007**. Version 2007. [s. l.]: Microsoft Corporation. 2007. 1 CD-ROM.

MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada: Uma Abordagem Aplicada**. Belo Horizonte: Editora UFMG, 2005, 295 p.

NORRIS, K. R. The bionomics of blowflies. **Ann. Rev. Entomol.**, [s.l.], v. 10, [s.n.] p.47-68, 1965. Disponível em: <https://www.annualreviews.org/doi/10.1146/annurev.en.10.010165.000403>. Acesso em: 15 ago. 2019.

NOVAIS, G. T.; BRITO, J. L. S. Tropical semiúmido e tropical semiseco: os tipos climáticos do domínio tropical brasileiro. *In*: Congresso Nacional de Geografia Física, 1, 2017, Campinas. **Anais [...]**, Campinas: Unicamp, 2017. p. 1411-1421. Disponível em: <https://pdfs.semanticscholar.org/acd3/403b761b71928630eff466efd46d7d7d917f.pdf>. Acesso em: 15 ago. 2019.

ODUM, E. P.; BARRETT, G. W. **Fundamentos de Ecologia**. 5ª ed. São Paulo: Cengage Learning, 2008. 632 p.

OLIVEIRA, H. V. **Dinâmica temporal da estrutura de comunidades de Drosophilidae (Insecta, Diptera) em frutos do Cerrado**. 2009. Dissertação (Mestrado em Ecologia)- Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, 2010. Disponível em: <https://repositorio.unb.br/handle/10482/4226>. Acesso em: 14 abr. 2020.

POPPE, J. L.; SCHMITZ, H. J.; VALENTE, V. L. S. The New World genus *Rhinoleucophenga* (Diptera: Drosophilidae): new species and notes on occurrence records. **Zootaxa**, [s.l.], v. 3955, n. 3, p. 349–370, 2015. Disponível em: https://www.researchgate.net/publication/276059386_The_New_World_genus_Rhinoleucophenga_Diptera_Drosophilidae_New_species_and_notes_on_occurrence_records. Acesso em: 12 abr. 2020.

PORTAL ACTION. **Testes de Normalidade**. c2020a. Disponível em: <http://www.portalaction.com.br/inferencia/testes-de-normalidade>. Acesso em: 04 fev. 2020.

PORTAL ACTION. **Diagnóstico de Homocedasticidade**. c2020b. Disponível em: <http://www.portalaction.com.br/analise-de-regressao/32-diagnostico-de-homocedasticidade>. Acesso em: 05 fev. 2020.

PORTAL ACTION. **Teste de igualdade das variâncias**. c2020c. Disponível em: <http://www.portalaction.com.br/anova/161-teste-de-igualdade-das-variancias>. Acesso em: 05 fev. 2020.

RAGA, A.; FILHO, M. F. S. Captura de *Zaprionus indianus* (Gupta) (DIP: Drosophilidae) em frascos de plásticos com iscas alimentares na cultura do figo. **Revista de Agricultura**, Piracicaba, v. 78, [s.n], p. 223-229, 2003. Disponível em: <http://revistadeagricultura.org.br/index.php/revistadeagricultura/article/view/296>. Acesso em: 15 mar. 2019.

R CORE TEAM. **R: A language and environment for statistical computing**. Version 3.6.1. Vienna, Austria: R Foundation for Statistical Computing. 2019. Disponível em: <https://www.R-project.org/>.

RIVERS, D. B.; THOMPSON, C.; BROGAN, R. Physiological trade-offs of forming maggot masses by necrophagous flies on vertebrate carrion. **Bulletin of Entomological Research**, [s.l.], v. 101, n. 5, p. 599–611, 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/21729395>. Acesso em: 04 abr. 2020.

ROHDE, C. *et al.* Espécies invasoras da família Drosophilidae (Diptera, insecta) em ambientes da caatinga de Pernambuco. **Anais da Academia Pernambucana de Ciência Agrônômica**, [s.l.], v. 7, [s.n], p.227-240, 2010. Disponível em: <http://www.journals.ufrpe.br/index.php/apca/article/view/140>. Acesso em: 15 mar. 2019.

ROQUE, F. **Drosophilídeos (Insecta, Diptera) da mata do Pitoco: diversidade e distribuição vertical**. 2009. Dissertação (Mestrado em Ecologia)- Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, 2009. Disponível em: <https://repositorio.unb.br/handle/10482/8279?mode=full>. Acesso em: 14 abr. 2020.

ROQUE, F.; MATA, R. A.; TIDON, R. Temporal and vertical drosophilid (Insecta; Diptera) assemblage fluctuations in a neotropical gallery Forest. **Biodivers Conserv**, [s.l.], v. 22, [s.n.], p. 657–672, 2013. Disponível em: <https://link.springer.com/article/10.1007/s10531-013-0433-4>. Acesso em: 20 mar. 2020.

SILVA, I. G. M. **Caracterização estrutural e ultraestrutural dos espermatozóides de representantes da família Drosophilidae (Diptera)**. 2013. Dissertação (Mestrado em Biologia Animal) - Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, 2013. Disponível em: <https://repositorio.unb.br/handle/10482/14400>. Acesso em: 20 mar. 2019.

SOKAL, R. R.; ROHLF, F. J. **Biometry**. 3ª ed. New York: W.H. Freeman and Company, 1995. 887 p.

TANTAWY, A.O. Effects of temperature on productivity and genetic variance of body size in populations of *Drosophila pseudoobscura*. **Genetics**, [s.l.], v. 46, [s.n.], p. 227-238, 1961. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/13775221>. Acesso em: 18 mar. 2019.

TANTAWY, A.O.; VETUKHIV, M.O. Effects of size on fecundity, longevity and viability in populations of *Drosophila pseudoobscura*. **The American Naturalist**, [s.l.], v. 94, [s.n.], p. 395-403, 1960. Disponível em: <https://www.journals.uchicago.edu/doi/pdfplus/10.1086/282143>. Acesso em: 18 mar. 2019.

THERNEAU, T. A Package for Survival Analysis in S. Version 2.38: Springer, New York. 2015. Disponível em: <https://CRAN.R-project.org/package=survival>.

TIDON, R.; LEITE, D. F.; LEÃO, B. F. D. Impact of the colonisation of *Zaprionus* (Diptera, Drosophilidae) in different ecosystems of the Neotropical Region: 2 years after the invasion. **Biological Conservation**, [s.l.], v. 112, [s.n.], p. 299–305, 2003. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0006320702003221>. Acesso em: 9 abr. 2020.

TORRES, F. R.; RAVAZZI, L. M. Seasonal variation in natural populations of *Drosophila* spp. (Diptera) in two woodlands in the State of São Paulo, Brazil. **Iheringia, Série Zoologia**, [s.l.], v. 96, n. 4, p. 437- 444, 2006. Disponível em:

http://www.scielo.br/scielo.php?pid=S0073-47212006000400008&script=sci_arttext. Acesso em: 19 mar. 2019.

UEL. **Teste de Shapiro-Wilk**. [201-?a]. Disponível em: <http://www.uel.br/projetos/experimental/pages/arquivos/Shapiro.html>. Acesso em: 04 fev. 2020.

UEL. **Teste de Levene**. [201-?b]. <http://www.uel.br/projetos/experimental/pages/arquivos/Levene.html>. Acesso em: 05 fev. 2020.

UFMG. **Distribuição Gaussiana**: Verificação da Suposição de Normalidade dos Dados. 2011. Disponível em: <https://dspaceprod01.grude.ufmg.br/dspace/bitstream/handle/OAUFMG/736/Aula%2012%20-%20Verifica%C3%A7%C3%A3o%20de%20Normalidade.pdf?sequence=5>. Acesso em: 04 fev. 2020.

VON ZUBEN, C. J. *et al.* Dynamics of a Mathematical Model of *Chrysomya megacephala* (Diptera: Calliphoridae). **J. Med. Entomol.**, [s.l.], v. 30, n. 2, p. 443-448, 1993. Disponível em: <https://academic.oup.com/jme/article-abstract/30/2/443/2221222?redirectedFrom=fulltext>. Acesso em: 18 ago. 2019.

VON ZUBEN, C. J. *et al.* Theoretical approaches to forensic entomology: I. Mathematical model of postfeeding larval dispersal. **J. Appl. Entomol.**, [s.l.], v. 120, n. 3, p. 379-382, 1996. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0418.1996.tb01623.x>. Acesso em: 18 ago. 2019.

VON ZUBEN, C. J.; STANGENHAUS, G.; GODOY, W. A. C. Competição larval em *Chrysomya megacephala* (F.) (Diptera: Calliphoridae): efeitos de diferentes níveis de agregação larval sobre estimativas de peso, fecundidade e investimento reprodutivo. **Rev. Bras. Biol.**, [s.l.], v. 60, n. 2, p. 195-203, 2000. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71082000000200002. Acesso em: 18 ago. 2019.

WELLS, J. D.; GREENBERG, B. Interaction between *Chrysomya rufifacies* and *Cochliomyia macellaria* (Diptera: Calliphoridae): the possible consequences of an invasion. **Bull. Entomol. Res.**, [s.l.], v. 82, [s.n.], p. 133-137, 1992. Disponível em: <https://www.cambridge.org/core/journals/bulletin-of-entomological-research/article/interaction-between-chrysomya-rufifacies-and-cochliomyia-macellaria-diptera-calliphoridae-the-possible-consequences-of-an-invasion/53436BBC73E6EC95338FA2C88F6113D4>. Acesso em: 18 ago. 2019.

ZAMPAULO, R. A. *et al.* Impactos em grutas graníticas na serra dos cocais (Valinhos-SP): patrimônio espeleológico desconhecido. *In:* Congresso Brasileiro de Espeleologia, 29., 2007, Ouro Preto. **Anais [...]**.Ouro Preto: Sociedade Brasileira de Espeleologia, 2007. p.335-340. Disponível em: http://www.sbe.com.br/anais29cbe/29cbe_335-340.pdf. Acesso em: 21 mar. 2020.

ZAR, J. H. **Biostatistical Analysis**. 4^a ed. Upper Saddle River: Prentice-Hall/Pearson, 1999. 663 p.

ZILLER, S. R.; ZALBA, S. Propostas de ação para prevenção e controle de espécies exóticas invasoras. **Natureza & Conservação**, [s.l.], v. 5, n. 2, p. 8-15, 2007. Disponível em: https://moodle.ufsc.br/pluginfile.php/822757/mod_resource/content/1/Ziller%20-%20Zalba%202%20-%20Natureza%20e%20Conservacao%202007%20Portugu%C3%AAs.pdf. Acesso em: 12 mar. 2019.

APÊNDICE A - TESTE DE BARTLETT

Um teste comumente utilizado para avaliar a homocedasticidade dos dados, um dos pressupostos para análises paramétricas, constitui-se o teste de Bartlett (B), o qual se apresenta bastante poderoso em dados com distribuição normal, sendo sensível a ausência de normalidade (PORTAL ACTION, c2020c; ZAR, 1999).

$$B = (N - k) \ln s_p^2 - \sum_{i=1}^k [(n_i - 1) \ln s_i^2] \quad (1.16)$$

Em que N corresponde ao tamanho amostral total; n_i as réplicas $j=1,2,...,n$ de determinado $i=1,2,...,k$ níveis do fator; enquanto os valores de s_p^2 e s_i^2 são fornecidos pelas seguintes equações:

$$s_p^2 = \frac{1}{N - k} \sum_{i=1}^k (n_i - 1) s_i^2 \quad (1.17)$$

$$s_i^2 = \sum_{j=1}^{n_i} \frac{(Y_{ij} - \bar{Y}_i)^2}{n_i - 1} \quad (1.18)$$

Sendo que Y_{ij} representa a observação de determinada réplica j e nível de fator i ; e $\bar{Y}_i$ a média de um determinado nível de fator (PORTAL ACTION, c2020c).

Quando o valor de B apresenta-se menor que o valor tabelado para determinado erro α e para o grau de liberdade $k-1$ da tabela qui-quadrado, ou, equivalentemente, quando a probabilidade do teste é maior que α , se aceita que as variâncias dos grupos são homogêneas ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$) (PORTAL ACTION, c2020c; ZAR, 1999).

APÊNDICE B - TESTE T-STUDENT

O teste t constitui-se de um método paramétrico padrão para testes de hipóteses de comparação de diferenças de duas médias de dados independentes ou relacionados, que apresentam como pressuposto normalidade dos dados (AYRES, 2011; SOKAL; ROHLF, 1995).

Considerando um teste de hipóteses referente às diferenças das médias de duas amostras normais, homocedásticas e independentes entre si, com médias μ_1 e μ_2 (sendo $H_0: \mu_1 - \mu_2 = 0$), a formula geral é representada abaixo:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \quad (1.19)$$

Sendo $\bar{X}_1$ e $\bar{X}_2$ as médias das amostras 1 e 2; n_1 e n_2 o tamanho amostral; e s_p a variância conjunta, calculada a partir da equação a seguir:

$$s_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \quad (1.20)$$

Em que s_1^2 e s_2^2 representam as variâncias das amostras 1 e 2 (AYRES, 2011; SOKAL; ROHLF, 1995).

Quando se pretende testar a hipótese de igualdade de médias provindas de duas amostras com distribuição normal, heterocedásticas e independentes entre si, calcula-se o valor de t da seguinte forma:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (1.21)$$

Para duas amostras relacionadas de mesmo tamanho amostral aplica-se o teste t de Student pareado (sendo $H_0: \mu_1 - \mu_2 = 0$), o qual pode ser calculado através da seguinte equação:

$$t = \frac{\bar{D}}{\left[\left(\sqrt{\frac{\sum (D_i - \bar{D})^2}{n - 1}} \right) \frac{1}{\sqrt{n}} \right]} \quad (1.22)$$

Sendo que D_i representa as diferenças entre os pares de observações; $\bar{D}$ a média destas diferenças; e n o tamanho amostral.

A partir do cálculo de t através das formulas acima, compara-se o valor obtido com aquele presente na tabela t -Student coluna $\alpha/2$ e linha $n_1 + n_2 - 2$ para um conjunto de dados homocedástico, ou $n_1 - 1$ ou $n_2 - 1$ (dependo do valor do menor n) para dados

heterocedásticos, ambos para o caso de duas amostras independentes. Em duas amostras relacionadas, o grau de liberdade, correspondente a linha, passa a apresentar o valor $n-1$. Quando t apresentar valores menores que $-t_{tab}$ ou maiores que t_{tab} , ou equivalentemente, quando a probabilidade do teste $p < \alpha$, rejeita-se a hipótese nula (H_0) (AYRES, 2011; SOKAL; ROHLF, 1995).

O teste t também pode ser utilizado para verificar os pares de médias, os quais apresentam diferenças significativas entre si, após a ANOVA rejeitar a hipótese nula sobre ausência de diferenças entre as médias das amostras. O cálculo da diferença mínima significativa (DMS) envolve o valor tabelado de t na tabela t -Student, linha α e coluna referente aos graus de liberdade (gl) do teste ANOVA ($t_{\alpha(gl)}$), o quadrado médio do erro da ANOVA ($QM_{dentro\ do\ grupos}$), e o número de repetições do experimento (r):

$$DMS = t_{\alpha(gl)} \sqrt{\frac{2QM_{dentro\ do\ grupos}}{r}} \quad (1.23)$$

O DMS obtido é comparado com os valores das diferenças das médias, em módulo, dos grupos, tomados 2 à 2. Quando o valor da diferença apresentar-se maior que o valor de DMS calculado, ou se possuir $p < \alpha$ considerado, se aceita a hipótese alternativa, na qual as diferenças entre as médias dos grupos são significativas (AYRES, 2011).

APÊNDICE C – ANOVA

ANOVA de um fator

Em análises mais simples, como a ANOVA de um fator ou “one-way”, são comparadas médias de grupos com base em um único tipo de tratamento, o qual pode apresentar subdividido em diversos níveis. Neste teste é calculada a soma dos quadrados totais (SQ_{total}), que pode ser subdividida em soma dos quadrados entre os grupos ($SQ_{entre\ grupos}$) e dentro dos grupos ($SQ_{dentro\ dos\ grupos}$), os quais representam a variação dos dados nos respectivos conjuntos considerados (GOTELLI; ELLISON, 2011; ZAR, 1999). Na ANOVA de um fator o SQ_{total} pode ser calculado pela seguinte formula:

$$SQ_{total} = \sum_{i=1}^a \sum_{j=1}^n (Y_{ij} - \bar{Y})^2 \quad (1.24)$$

Sendo que Y_{ij} corresponde às observações de um determinado tratamento $i=1,2,...,a$ e réplica $j=1,2,...,n$; e $\bar{Y}$ a média geral (GOTELLI; ELLISON, 2011).

As $SQ_{entre\ grupos}$ e $SQ_{dentro\ dos\ grupos}$ (variação residual ou do erro) são calculados através das respectivas equações:

$$SQ_{entre\ grupos} = \sum_{i=1}^a (\bar{Y}_i - \bar{Y})^2 \quad (1.25)$$

$$SQ_{dentro\ dos\ grupos} = \sum_{i=1}^a \left[\sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2 \right] \quad (1.26)$$

Em que $\bar{Y}_i$ representa a média de cada tratamento $i=1,2,...,a$.

A partir dos valores das somas dos quadrados, é possível calcular o valor dos quadrados médios entre grupos ($QM_{entre\ grupos}$) e dentro dos grupos ($QM_{dentro\ do\ grupo}$), os quais consistem em estimativas das variâncias (σ^2) dos diferentes níveis de tratamentos, por meio da divisão pelos seus respectivos graus de liberdade $a-1$ e $a(n-1)$. O cálculo da razão F , constitui-se na divisão entre $QM_{entre\ grupos}$ por $QM_{dentro\ do\ grupo}$, sendo que resultado deve ser comparado com o valor de F tabelado para numerador $a-1$, denominador $a(n-1)$ e determinado erro α do teste, ou verificado o valor da probabilidade do teste em relação ao seu nível α considerado para aceitar ou, de maneira equivalente, rejeitar a hipótese de igualdade entre médias dos grupos (AYRES, 2011; GOTELLI; ELLISON, 2011; ZAR, 1999).

A aceitação da hipótese alternativa (H_a) sobre a influência do tratamento aplicado nas médias dos grupos amostrais, geralmente ocorre quando a diferença dos quadrados médios entre os grupos apresenta-se significativamente maior que aquele dentro dos grupos (GOTELLI; ELLISON, 2011; ZAR, 1999).

ANOVA de dois fatores

A ANOVA de dois fatores analisa as amostras expostas a dois diferentes tratamentos e seus níveis, avaliando os efeitos dos fatores individualmente e de sua interação. Semelhantemente a ANOVA de um fator, calcula-se a soma dos quadrados totais (SQ_{total}), entre grupos ($SQ_{entre\ grupos}$) e dentro dos grupos ($SQ_{dentro\ dos\ grupos}$), porém os grupos são substituídos por células ($SQ_{células}$ e $SQ_{dentro\ de\ células}$, respectivamente). Em que cada célula (Y_{ij}) consiste de uma combinação dos diferentes níveis de primeiro fator (A , com $i=1,2,...,a$ níveis) com o segundo fator (B , com $j=1,2,...,b$ níveis), podendo possuir uma ou mais réplicas ($k=1,2,...,n$ réplicas) por célula (GOTELLI; ELLISON, 2011; ZAR, 1999).

A soma dos quadrados do primeiro fator (SQ_A) é calculada pela estimativa de sua média para todos os níveis do segundo tratamento, enquanto o segundo fator (SQ_B) para todos os níveis do primeiro. As soma dos quadrados do primeiro (SQ_A) e do segundo fator (SQ_B) são obtidas pelas seguintes equações:

$$SQ_A = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_i - \bar{Y})^2 \quad (1.27)$$

$$SQ_B = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_j - \bar{Y})^2 \quad (1.28)$$

Em que $\bar{Y}_i$ constitui-se da média do primeiro fator; $\bar{Y}_j$ a média do segundo fator; e $\bar{Y}$ a média geral (GOTELLI; ELLISON, 2011).

Além disso, calcula-se a soma dos quadrados da interação dos dois fatores (SQ_{AB}):

$$SQ_{AB} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y})^2 \quad (1.29)$$

Sendo que $\bar{Y}_{ij}$ representa a média da célula com uma determinada combinação do nível do primeiro fator com o segundo, podendo apresentar $k=1,2,...,n$ réplicas.

A partir dos valores das somas dos quadrados, pode-se calcular os quadrados médios do fator A (QM_A), B (QM_B), interação A e B (QM_{AB}) e dentro das células ($QM_{dentro\ de\ células}$) ao

dividi-los por seus respectivos graus de liberdade $a-1$, $b-1$, $(a-1)(b-1)$ e $ab(n-1)$. Calculados os valores dos quadrados médios obtêm-se três razões F pela razão de $QM_A / QM_{\text{dentro de células}}$, $QM_B / QM_{\text{dentro de células}}$, e $QM_{AB} / QM_{\text{dentro de células}}$, verificando-se a significância do efeitos dos fatores e de sua interação nas observações, a partir dos valores F obtidos, comparando-os com aqueles tabelados para determinado nível α do teste, numerador $a-1$, $b-1$ ou $(a-1)(b-1)$, dependendo do fator ou da interação testada, e denominador $ab(n-1)$, ou através do valor da probabilidade do teste. Comumente, o efeito da interação entre os fatores é considerado insignificante, quando os efeitos dos fatores analisados individualmente apresentam valores altamente significantes (GOTELLI; ELLISON, 2011; ZAR, 1999).

ANOVA multifatorial

Quando estudo envolve mais de dois fatores, utiliza-se a ANOVA multifatorial para análise da influência individual dos fatores e de suas interações na variável reposta, por meio da combinação de tratamentos diversos e de cada um de seus níveis, tanto em delineamentos que envolvem blocos aleatorizados, repetições e subamostragem (GOTELLI; ELLISON, 2011; ZAR, 1999).

Em geral, os cálculos das somas dos quadrados são semelhantes aos delineamentos de ANOVA anteriores, diferenciando pelos graus de liberdade utilizados para calcular os quadrados médios e consequentemente as razões F . Porém, à medida que se eleva a quantidade de fatores testados e de seus níveis, aumenta-se o número de combinações distintas possíveis e consequentemente a complexidade da análise, a qual geralmente é realizada por meio de programas computacionais estatísticos (ZAR, 1999).

No delineamento mais simples de três fatores, com primeiro fator A com $i=1,2,...,a$ níveis, o segundo fator B com $j=1,2,...,b$ níveis, e o terceiro fator C com $k=1,2,...,c$ níveis, havendo n réplicas para cada combinação distinta de cada um dos níveis dos três fatores, totalizando um total de $i \times j \times k \times n$ observações. No delineamento considerado calculam-se as somas dos quadrados dos fatores individualmente, de suas interações e dos erros. Para cálculo do quadrado médio dos fatores, divide-se as somas dos quadrados dos fatores pelos graus de liberdade referentes ao número de níveis subtraído o valor 1. Para os quadrados médios das interações, multiplicam-se os graus de liberdade dos respectivos fatores, e utilizam-se os valores obtidos para dividir as somas dos quadrados das interações. Para cálculo do quadrado médio do erro, subtraí os valores dos graus de liberdade dos fatores e das interações do total $(N-1)$, posteriormente utilizando o valor obtido para dividir a soma do quadrado do erro.

Desta forma, obtém setes razões F , pela divisão dos quadrados médios dos fatores individuais e de suas interações pelo erro. Estas razões testam as hipóteses nulas de igualdade entre as médias e consequentemente a ausência de efeitos dos três fatores individualmente, assim como hipóteses acerca da independência dos efeitos dos fatores, referindo-se a influência significativa de suas interações ($A \times B$, $A \times C$, $B \times C$ e $A \times B \times C$) (ZAR, 1999).

APÊNDICE D- TRANSFORMAÇÃO DE DADOS

Nem sempre o conjunto de dados estudados (Y) está em conformidade com os pressupostos de normalidade, homocedasticidade, aditividade e aleatoriedade dos resíduos para utilização de testes paramétricos, sendo necessária a anterior aplicação de transformações simultâneas de dados (GOTELLI; ELLISON, 2011; ZAR, 1999). Estas transformações usualmente normalizam a distribuição dos valores e uniformizam as variâncias, simplificando funções complexas (GOTELLI; ELLISON, 2011).

Dentre estas transformações, uma das mais comumente utilizadas consiste na transformação logarítmica ($Y' = \log_b(Y)$), a qual geralmente torna fatores multiplicativos em aditivos, assim como homogeneíza médias e variâncias, aproximando os dados de uma distribuição normal. Em geral, a conversão dos dados em qualquer base ($b=2, e, 10...$) produz um resultado satisfatório, sendo que geralmente em estudos biológicos utiliza-se a base logarítmica 10, somando 1 ao valor da observação original, buscando evitar resultados negativos (GOTELLI; ELLISON, 2011; ZAR, 1999).

Outra transformação utilizada é a raiz quadrada ($Y' = \sqrt{Y}$), a qual é aplicada, principalmente para dados de contagem, ou dados que possuam variâncias proporcionais as médias, gerando, muitas vezes, resultados com distribuição normal e variâncias independentes da média (GOTELLI; ELLISON, 2011; ZAR, 1999).

A transformação arcoseno ($Y' = \arcsin\sqrt{Y}$) é usualmente aplicada para dados com distribuições binominais, como por exemplo, porcentagens e proporções. Este tipo de conversão remove a dependência da variância em relação à média, gerando dados em radianos, porém não é muito eficiente com observações extremas (0 e 100%) (GOTELLI; ELLISON, 2011; ZAR, 1999).

A transformação inversa ($Y' = 1/Y$), é constantemente aplicada a dados sobre taxas ($aXY=1$, em que a consiste de uma constante, X e Y as variáveis), que apresentam distribuição gráfica hiperbólica, sendo convertidas em relações lineares ($1/Y=aX$), mais simplificadas para serem analisadas (GOTELLI; ELLISON, 2011).

APÊNDICE E- CORRELAÇÃO

Correlação de Pearson

O coeficiente de correlação de Pearson pode ser calculado através da seguinte equação:

$$r = \frac{\sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sqrt{(\sum (X_i - \bar{X})^2) (\sum (Y_i - \bar{Y})^2)}} \quad (1.30)$$

Em que X_i e $\bar{X}$ representam respectivamente os valores das observações e da média da primeira variável; e Y_i e $\bar{Y}$ as observações e a média da segunda variável, ambos de uma amostra de tamanho n . Semelhantemente a correlação de Spearman, o coeficiente de Pearson varia entre $-1 \leq r \leq +1$, possibilitando a mesma interpretação sobre a intensidade e direção da correlação (AYRES, 2011; ZAR, 1999).

Além do cálculo do valor do coeficiente r , pode testar hipóteses quanto sua significância através da aplicação da estatística t ou F :

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad (1.31)$$

$$F = \frac{1+|r|}{1-|r|} \quad (1.32)$$

Neste caso, a hipótese nula sobre a não existência de correlação ($\rho = 0$) é aceita quando os valores de t e F calculados forem $-t_{tab} \leq t \leq t_{tab}$ (t tabelado na coluna $\alpha/2$ e linha $n-2$ da tabela t -Student) e $-F_{tab} \leq F \leq F_{tab}$ (F tabelado na coluna α , com numerador e denominador $n-2$ da tabela F) respectivamente, ou as probabilidades dos testes forem $p \geq \alpha$ (AYRES, 2011; ZAR, 1999).

Correlação de Kendall

Outra forma de estimar a associação entre as variáveis constitui-se na correlação não-paramétrica de Kendall (τ), na qual se atribuem postos as observações das variáveis (AYRES, 2011).

$$\tau = \frac{S}{\frac{1}{2}N(N+1)} \quad (1.33)$$

Sendo que S corresponde a soma dos postos; e N o tamanho amostral.

Semelhantemente aos testes anteriores, é possível verificar a significância do parâmetro populacional ρ , ao converter o coeficiente τ em z , como apresentado na seguinte equação:

$$z = \frac{\tau}{\sqrt{\frac{2(2N+5)}{9N(N-1)}}} \quad (1.34)$$

A aceitação da hipótese nula sobre a ausência de correlação ($\rho=0$) entre as variáveis ocorre quando a probabilidade do teste $p > \alpha$ (AYRES, 2011).

Coeficiente de Contingência

Quando ambas as variáveis são nominais dicotômicas (1=presença, 0=ausência) ou ordinais, os dados podem ser organizados em tabelas de contingências 2×2 . Nestas tabelas, uma das variáveis está disposta nas linhas, enquanto a outra nas colunas, sendo que cada célula corresponde a combinação dos seus níveis, presença ou ausência. As margens da tabela de contingência correspondem a soma dos valores de cada linha ($i=1, 2, \dots, l$) e cada coluna ($j=1, 2, \dots, c$) (AYRES, 2011; ZAR, 1999).

A partir da tabela de contingência, calcula-se o valor de qui-quadrado (X^2) para os dados, pela diferença dos valores observados (O_{ij}) em relação aos esperados (E_j), os quais são estimados pelos totais marginais (AYRES, 2011).

$$X^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_j)^2}{E_j} \quad (1.35)$$

Depois de calculado o valor de X^2 , pode-se substituí-lo na fórmula a seguir para estimar o coeficiente de contingência (C), assim como pode ser utilizado para encontrar a probabilidade do teste, rejeitando a ausência de correlação quando $p < \alpha$ (AYRES, 2011).

$$C = \sqrt{\frac{X^2}{N + X^2}} \quad (1.36)$$

Sendo que N representa a soma do número de categorias da tabela de contingência. Quando o valor de C é próximo a zero, ocorre uma baixa associação entre as variáveis, enquanto que, o valor próximo a 1 indica uma alta associação (AYRES, 2011).

Coefficiente de Concordância

O coeficiente de concordância (r_c) é comumente utilizado para comparar duas amostras que se referem ao desempenho de um método, equipamento ou profissional.

$$r_c = \frac{2 \sum (X_i - \bar{X}) (Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2 + \sum (Y_i - \bar{Y})^2 + (n - 1)(\bar{X} - \bar{Y})} \quad (1.37)$$

Sendo que X_i e $\bar{X}$ correspondem respectivamente os valores das observações e da média da primeira variável; Y_i e $\bar{Y}$ as observações e a média da segunda variável; e n o tamanho da amostra (AYRES, 2011; ZAR, 1999).

O coeficiente de concordância varia entre $-1 \leq r_c \leq +1$, e apresenta interpretação da intensidade e direção da correlação das variáveis semelhante ao coeficiente de Pearson, porém geralmente apresenta um valor menor que o último (AYRES, 2011; ZAR, 1999).

APÊNDICE F- REGRESSÃO SIMPLES E MÚLTIPLA

A análise de regressão permite determinar a intensidade de dependência da variável (Y) resposta em relação às variáveis preditoras ($X_1, X_2, \dots, X_i$), gerando um modelo que melhor descreve a relação entre as variáveis. Desta maneira, a análise de regressão possibilita a previsão de valores da variável dependente a partir de uma ou mais preditoras de interesse, sendo muito utilizada em estudos biológicos sobre modelagem de distribuição de espécies ao longo do tempo (AYRES, 2011; ZAR, 1999).

Antes de optar pela utilização de um modelo de regressão, é necessário verificar se a variável dependente Y apresenta distribuição normal e variância constante, assim como a independência entre os valores desta variável (AYRES, 2011).

Regressão linear simples

O modelo mais simples de regressão constitui-se no modelo linear, em que a relação entre as variáveis é adequada aos elementos da equação da reta:

$$\hat{Y}_i = a + bX_i + \varepsilon \quad (1.38)$$

Em que $\hat{Y}_i$ representa o valor da variável resposta previsto pela regressão linear; X_i a variável preditora; a e b os coeficientes de intercepto e inclinação, e ε o resíduo (AYRES, 2011; ZAR, 1999).

Os resíduos constituem-se de desvios dos valores observados para Y_i em relação valor predito pela reta de regressão ($\hat{Y}_i$). Desta forma, a reta gerada pelo modelo de regressão linear simples é aquela que melhor representa os dados, por apresentar o menor valor referente à soma dos quadrados dos desvios (AYRES, 2011; ZAR, 1999).

O coeficiente de inclinação ou regressão (b), na regressão linear simples, representa a declividade da reta e as relações entre as variáveis preditora e resposta, como a ausência de associação ($b=0$), associação positiva ($b>0$) e negativa ($b<0$). O coeficiente b pode ser calculado da seguinte forma:

$$b = \frac{\sum(X_i - \bar{X}_i)(Y_i - \bar{Y}_i)}{\sum(X_i - \bar{X}_i)^2} \quad (1.39)$$

Sendo que X_i e $\bar{X}_i$ o valor amostral e a média da variável independente; e $\bar{Y}_i$ a média da variável dependente (AYRES, 2011; ZAR, 1999).

O coeficiente de intercepto (a), na regressão linear simples, consiste no valor onde a reta atravessa o eixo vertical da coordenada na representação gráfica, podendo ser obtido pela seguinte equação:

$$a = \bar{Y}_i - b\bar{X}_i \quad (1.40)$$

Após o cálculo dos componentes da reta de regressão, podem-se realizar testes de hipóteses do tipo ANOVA para averiguar a significância dos coeficientes de intercepto e inclinação ($H_0: a=0$; $H_0: b=0$), assim como para comparar os coeficientes de duas regressões ($H_0: a_1=a_2$; $H_0: b_1=b_2$). Pode-se também calcular o coeficiente de explicação ou determinação (R^2), que varia de 0 à 1, e expressa o nível da explanação da distribuição da variável resposta pela variável preditora, sendo que $R^2=1$, constitui-se de uma explicação perfeita dos dados (AYRES, 2011; ZAR, 1999).

Regressão múltipla

A regressão múltipla generaliza a regressão linear simples, apresentando uma variável resposta determinada por várias variáveis preditoras ($X_1, X_2, X_3, \dots, X_i$), cujos valores são geralmente organizados em matrizes para facilitar os cálculos computacionais (AYRES, 2011; ZAR, 1999). A formula geral da regressão múltipla é apresentada a seguir, onde é possível verificar a existência de um único coeficiente de intercepto (a), e vários coeficientes de regressão (b) acompanhando as variáveis independentes ($X_1, X_2, \dots, X_i$):

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 \dots + b_iX_i \quad (1.41)$$

No caso da regressão múltipla, b representa o coeficiente parcial de regressão, o qual mede o efeito da alteração da variável dependente quando se altera em uma unidade de certa variável independente, enquanto as outras são mantidas constantes (ZAR, 1999).

Semelhantemente a regressão linear simples, na regressão múltipla pode-se também calcular as significâncias dos coeficientes a e b s ($H_0: a=0$; $H_0: b_1=b_2=\dots b_i=0$), comparar coeficientes de dois ou mais modelos de regressão múltiplas, e estimar o coeficiente de determinação (R^2) para cada uma das variáveis independentes, verificando quais explicam melhor a distribuição das medidas da variável dependente (AYRES, 2011; ZAR, 1999).

A partir do coeficiente de determinação da regressão múltipla, é possível calcular o coeficiente de correlação múltipla ao extrair a raiz quadrada positiva de R^2 (ZAR, 1999).

Regressões polinomiais

Quando os dados não se adequam aos modelos de regressão abordados anteriormente, pode-se optar pela utilização de regressões polinomiais, cujas equações geram gráficos não lineares (parábolas, hipérboles, entre outros) (AYRES, 2011).

Em geral, as regressões polinomiais se caracterizam por apresentar apenas uma variável resposta e uma preditora, sendo que a última apresenta-se acompanhada por coeficientes de regressão distintos ($b_1, b_2, b_3... b_k$), além de serem elevadas a diferentes potências dependendo do modelo considerado (AYRES, 2011). A seguir são apresentadas, respectivamente, as equações gerais das regressões polinomiais quadrática, cúbica e quártica:

$$\hat{Y} = a + b_1X + b_2X^2 \quad (1.42)$$

$$\hat{Y} = a + b_1X + b_2X^2 + b_3X^3 \quad (1.43)$$

$$\hat{Y} = a + b_1X + b_2X^2 + b_3X^3 + b_4X^4 \quad (1.44)$$

Existem outros modelos de regressões não lineares, porém não serão abordados neste texto, por apresentar maior enfoque na regressão linear simples e múltipla.

APÊNDICE G- ANÁLISE DE SOBREVIVÊNCIA

Atuarial

O modelo Atuarial foi originalmente criado no século XX por demógrafos, sendo comumente aplicado em uma amostra de tamanho n , apresenta intervalos de coletas de dados constantes e longos, delimitados por tempos t_j (sendo $j=1, 2, \dots, s$) que nem sempre representam os tempos de falhas exatos (AYRES, 2011; COLOSIMO; GIOLO, 2006). O estimador atuarial ($\hat{S}(t)$) da sobrevivência acumulada dos indivíduos de uma amostra, pode ser obtido através da seguinte formula:

$$\hat{S}(t) = \prod_{j=1}^s (1 - \hat{q}_{j-1}) \quad (1.45)$$

Sendo $\hat{q}_j$ a probabilidade de ocorrência de falhas em um determinado intervalo t_{j-1} e t_j , podendo ser calculado através da seguinte equação:

$$\hat{q}_j = \frac{d_j}{n_j - \frac{w_j}{2}} \quad (1.46)$$

Sendo d_j o número de falhas (reincidências ou óbitos); w_j o número de censuras, em que ambos referem-se ao intervalo de t_{j-1} até t_j ; n_j indivíduos sobreviventes em t_{j-1} (AYRES, 2011; COLOSIMO; GIOLO, 2006).

Devido a sua origem e a utilização de intervalos arbitrários longos, o modelo Atuarial não fornece estimativas tão confiáveis em estudos clínicos quando comparados com modelos Kaplan-Meier ou Nelson-Aalen, devido ao reduzido tamanho das amostras associados a estes tipos de pesquisas (COLOSIMO; GIOLO, 2006).

Nelson-Aalen

O estimador de sobrevivência acumulada de Nelson-Aalen ($\tilde{S}(t)$) apresenta propriedades semelhantes ao modelo Kaplan-Meier, obtendo-se resultados maiores ou idênticos (COLOSIMO; GIOLO, 2006).

$$\tilde{S}(t) = \exp\{-\tilde{\Lambda}_{NA}(t)\} \quad (1.47)$$

Em que $\exp$ representa a função exponencial do número natural e ; e $\tilde{\Lambda}_{NA}(t)$ o estimador de risco acumulado de Nelson-Aalen, calculado da seguinte maneira:

$$\tilde{\Lambda}_{NA}(t) = \sum_{j; t_j < t} \left(\frac{d_j}{n_j} \right) \quad (1.48)$$

Sendo d_j o número de falhas em t_j ; e n_j o número sobreviventes em t_{j-1} (COLOSIMO; GIOLO, 2006).

Modelo de Risco proporcional de Cox

O modelo de Cox é bastante utilizado em estudos clínicos, pois se constitui de um modelo flexível que permite verificar a influência dos tratamentos nos tempos de falha dos pacientes através de ajustes de covariáveis (idade, sexo, entre outros) a um modelo de regressão. Por consistir de um modelo semi-paramétrico, este não pressupõe homocedasticidade e distribuição normal de dados, por outro lado, exige que as razões das taxas de falhas ($\lambda(t)$) sejam proporcionais entre diferentes pacientes, independentemente do tempo considerado ($\lambda_1(t)/\lambda_0(t) = K$, em que K representa uma constante) (AYRES, 2011; COLOSIMO; GIOLO, 2006).

O modelo de Cox pode ser obtido pela aplicação do método de máxima verossimilhança parcial, sendo representado simplificadaamente pela seguinte equação:

$$\lambda(t) = \lambda_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p) \quad (1.49)$$

Sendo $\lambda(t)$ a taxa de falha do segundo grupo para um determinado intervalo de tempo, considerando um total de dois grupos, o controle e tratamento; $\lambda_0(t)$ a taxa de falha do primeiro grupo; $\exp$ a função exponencial do número natural e ; p a covariável em questão; β coeficiente que maximiza a função de verossimilhança parcial, sendo associado às covariáveis e responsável por estimar seus efeitos sobre a função; e x o valor das covariáveis para os diferentes pacientes e seus respectivos tempos de falha (COLOSIMO; GIOLO, 2006).

Quando o coeficiente β assume um valor positivo, representa que uma elevação dos valores das covariáveis associadas, aumentaria a taxa de risco, reduzindo a sobrevivência dos pacientes. Enquanto um $\beta < 0$, representa um decréscimo da taxa de risco pela elevação do valor da covariável associada ao coeficiente de máxima verossimilhança parcial (AYRES, 2011; COLOSIMO; GIOLO, 2006).

APÊNDICE H- ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais constitui-se de um tipo de análise multivariada que resume o número de variáveis ($X_1, X_2, \dots, X_p$) estudadas em índices gerados a partir da combinação linear destas, conhecidos como componentes principais (Z_i). Este tipo de análise é comumente utilizado em estudos biológicos onde existem muitas variáveis correlacionadas, como por exemplo, medidas de diferentes partes do corpo de um ser vivo (MANLY, 2008; MINGOTI, 2005).

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (1.50)$$

Em que a_{ip} referem-se às constantes que acompanham as variáveis, que se constituem de autovetores obtidos a partir dos autovalores ou coeficientes de variância de cada componente principal λ_p , calculados por sua vez através da matriz de covariâncias do conjunto de dados referentes aos valores $i=1,2,\dots,n$ de cada uma das p variáveis (MANLY, 2008; MINGOTI, 2005).

Uma matriz de dados de i linhas e p variáveis (colunas) podem fornecer até p componentes principais não correlacionadas entre si ($\lambda_1 + \lambda_2 + \dots + \lambda_p = p$), as quais apresentam diferentes dimensões, relações e representações da variabilidade de cada uma das variáveis. Desta forma, as primeiras componentes principais explicam melhor a variabilidade dos dados devido aos valores mais elevados de λ_p ($\lambda_1 > \lambda_2 > \dots > \lambda_p$), sendo, geralmente, aquelas mais interessantes de serem estudadas (MANLY, 2008; MINGOTI, 2005).

Após o cálculo dos autovalores e autovetores das componentes principais, escolhe-se aquelas que explicam melhor a variância dos dados, ou seja, interessantes para o fenômeno estudado, substituindo-se os valores originais das variáveis e calculando-se escores para as componentes escolhidos. A partir destes escores, é possível interpretar melhor a dimensão que as componentes explicam ao colocá-las uma contra a outra em um gráfico de dispersão, verificando-se tendências nas distribuições dos pontos (MANLY, 2008; MINGOTI, 2005).

A análise de componentes principais tem como pressuposto a existência de elevada correlação (positiva ou negativa) entre as variáveis estudadas. Além disso, é possível reduzir a dominância de uma variável com maior variância nas componentes principais, ao realizar a padronização das medidas das variáveis para $\bar{X} = 0$ e $s^2 = 1$, por meio da subtração da média em relação à observação, seguida da divisão pelo seu desvio padrão. A partir da matriz de valores padronizados, calculam-se os autovalores e autovetores através da matriz de correlação entre as variáveis (MANLY, 2008; MINGOTI, 2005).

APÊNDICE I- RESULTADOS DA ANÁLISE DE SOBREVIVÊNCIA DO MUNICÍPIO BEBEDOURO, SP

Tabela 21- Resultados da análise de sobrevivência para a população feminina de Bebedouro.

(continua)						
Tempos de falhas (t_j)	Sobreviventes em t_j	Falhas (d_j)	Probabilidades de falhas em t_j ($\hat{q}_j$)	Taxas de falhas entre t_{j-1} e t_j	Sobrevivências acumuladas	Intervalo de confiança de 95%
20	180	1	0,006	0,000	0,994	0,984 a 1,0000
23	179	1	0,006	0,002	0,989	0,974 a 1,0000
27	178	1	0,006	0,001	0,983	0,965 a 1,0000
32	177	1	0,006	0,001	0,978	0,956 a 0,999
33	176	1	0,006	0,006	0,972	0,948 a 0,996
35	175	1	0,006	0,003	0,967	0,940 a 0,993
36	174	1	0,006	0,006	0,961	0,933 a 0,989
37	173	1	0,006	0,006	0,956	0,925 a 0,986
38	172	1	0,006	0,006	0,950	0,918 a 0,982
40	171	1	0,006	0,003	0,944	0,911 a 0,978
41	170	1	0,006	0,006	0,939	0,904 a 0,974
44	169	1	0,006	0,002	0,933	0,897 a 0,970
46	168	1	0,006	0,003	0,928	0,890 a 0,966
51	167	3	0,018	0,004	0,911	0,869 a 0,953
52	164	2	0,012	0,012	0,900	0,856 a 0,944
53	162	3	0,019	0,019	0,883	0,836 a 0,930
55	159	1	0,006	0,003	0,878	0,830 a 0,926
56	158	1	0,006	0,006	0,872	0,823 a 0,921
57	157	5	0,032	0,032	0,844	0,791 a 0,897
58	152	1	0,007	0,007	0,839	0,785 a 0,893
59	151	1	0,007	0,007	0,833	0,779 a 0,888
60	150	3	0,020	0,020	0,817	0,760 a 0,873
61	147	3	0,020	0,020	0,800	0,742 a 0,858
62	144	3	0,021	0,021	0,783	0,723 a 0,843
63	141	1	0,007	0,007	0,778	0,717 a 0,838
64	140	2	0,014	0,014	0,767	0,705 a 0,828
65	138	1	0,007	0,007	0,761	0,699 a 0,823
66	137	1	0,007	0,007	0,756	0,693 a 0,818
67	136	1	0,007	0,007	0,750	0,687 a 0,813
68	135	3	0,022	0,022	0,733	0,669 a 0,798
69	132	2	0,015	0,015	0,722	0,657 a 0,788
70	130	2	0,015	0,015	0,711	0,645 a 0,777
71	128	8	0,063	0,063	0,667	0,598 a 0,735
72	120	7	0,058	0,058	0,628	0,557 a 0,700
73	113	3	0,027	0,027	0,611	0,540 a 0,682
74	110	1	0,009	0,009	0,606	0,534 a 0,677

Tabela 21- Resultados da análise de sobrevivência para a população feminina de Bebedouro.

(conclusão)						
Tempos de falhas (t_j)	Sobreviventes em t_j	Falhas (d_j)	Probabilidades de falhas em t_j ($\hat{q}_j$)	Taxas de falhas entre t_{j-1} e t_j	Sobrevivências acumuladas	Intervalo de confiança de 95%
75	109	4	0,037	0,037	0,583	0,511 a 0,655
76	105	6	0,057	0,057	0,550	0,477 a 0,623
77	99	5	0,051	0,051	0,522	0,449 a 0,595
78	94	3	0,032	0,032	0,506	0,432 a 0,579
79	91	6	0,066	0,066	0,472	0,399 a 0,545
80	85	5	0,059	0,059	0,444	0,372 a 0,517
81	80	2	0,025	0,025	0,433	0,361 a 0,506
82	78	3	0,039	0,038	0,417	0,345 a 0,489
83	75	5	0,067	0,067	0,389	0,318 a 0,460
84	70	5	0,071	0,071	0,361	0,291 a 0,431
85	65	7	0,108	0,108	0,322	0,254 a 0,390
86	58	6	0,103	0,103	0,289	0,223 a 0,355
87	52	5	0,096	0,096	0,261	0,197 a 0,325
88	47	2	0,043	0,043	0,250	0,187 a 0,313
89	45	2	0,044	0,044	0,239	0,177 a 0,301
90	43	7	0,163	0,163	0,200	0,142 a 0,258
91	36	6	0,167	0,167	0,167	0,112 a 0,221
92	30	6	0,200	0,200	0,133	0,084 a 0,183
93	24	4	0,167	0,167	0,111	0,065 a 0,157
94	20	2	0,100	0,100	0,100	0,056 a 0,144
95	18	6	0,333	0,333	0,067	0,030 a 0,103
96	12	4	0,333	0,333	0,044	0,014 a 0,075
97	8	3	0,375	0,375	0,028	0,004 a 0,052
98	5	2	0,400	0,400	0,017	0,000 a 0,035
99	3	1	0,333	0,333	0,011	0,000 a 0,026
101	2	1	0,500	0,250	0,006	0,000 a 0,016
104	1	1	1,000	0,333	0,000	0,000 a 0,000

Fonte: dados da análise

Tabela 22- Resultados da análise de sobrevivência para a população masculina de Bebedouro.

(continua)						
Tempos de falhas (t_j)	Sobreviventes em t_j	Falhas (d_j)	Probabilidades de falhas em t_j ($\hat{q}_j$)	Taxas de falhas entre t_{j-1} e t_j	Sobrevivências acumuladas	Intervalo de confiança de 95%
0,08	205	1	0,005	0,061	0,995	0,986 a 1,000
12	204	1	0,005	0,000	0,990	0,977 a 1,000
18	203	2	0,010	0,002	0,981	0,962 a 0,999
22	201	1	0,005	0,001	0,976	0,954 a 0,997
26	200	2	0,010	0,003	0,966	0,941 a 0,991
27	198	1	0,005	0,005	0,961	0,934 a 0,987

Tabela 22- Resultados da análise de sobrevivência para a população masculina de Bebedouro.

(continua)

Tempos de falhas (t_j)	Sobreviventes em t_j	Falhas (d_j)	Probabilidades de falhas em t_j ($\hat{q}_j$)	Taxas de falhas entre t_{j-1} e t_j	Sobrevivências acumuladas	Intervalo de confiança de 95%
28	197	1	0,005	0,005	0,956	0,928 a 0,984
29	196	4	0,020	0,020	0,937	0,903 a 0,970
33	192	2	0,010	0,003	0,927	0,891 a 0,962
34	190	2	0,011	0,011	0,917	0,879 a 0,955
36	188	1	0,005	0,003	0,912	0,873 a 0,951
39	187	2	0,011	0,004	0,902	0,862 a 0,943
41	185	2	0,011	0,005	0,893	0,851 a 0,935
43	183	3	0,016	0,008	0,878	0,833 a 0,923
44	180	1	0,006	0,006	0,873	0,828 a 0,919
45	179	1	0,006	0,006	0,868	0,822 a 0,915
46	178	3	0,017	0,017	0,854	0,805 a 0,902
48	175	1	0,006	0,003	0,849	0,780 a 0,898
49	174	2	0,012	0,011	0,839	0,789 a 0,890
50	172	3	0,017	0,017	0,824	0,772 a 0,876
51	169	2	0,012	0,012	0,815	0,761 a 0,868
52	167	1	0,006	0,006	0,810	0,756 a 0,863
53	166	2	0,012	0,012	0,800	0,745 a 0,855
54	164	3	0,018	0,018	0,785	0,729 a 0,842
55	161	4	0,025	0,025	0,766	0,708 a 0,824
56	157	4	0,026	0,025	0,746	0,687 a 0,806
57	153	4	0,026	0,026	0,727	0,666 a 0,788
58	149	8	0,054	0,054	0,688	0,624 a 0,751
60	141	5	0,036	0,018	0,663	0,599 a 0,728
61	136	5	0,037	0,037	0,639	0,573 a 0,705
62	131	2	0,015	0,015	0,629	0,563 a 0,695
63	129	4	0,031	0,031	0,610	0,543 a 0,676
64	125	3	0,024	0,024	0,595	0,528 a 0,662
65	122	3	0,025	0,025	0,581	0,513 a 0,648
66	119	3	0,025	0,025	0,566	0,498 a 0,634
67	116	4	0,035	0,034	0,546	0,478 a 0,614
68	112	5	0,045	0,045	0,522	0,454 a 0,590
69	107	6	0,056	0,056	0,493	0,424 a 0,561
70	101	4	0,040	0,040	0,473	0,405 a 0,541
71	97	8	0,083	0,082	0,434	0,366 a 0,502
72	89	2	0,023	0,022	0,424	0,357 a 0,492
73	87	3	0,035	0,034	0,410	0,342 a 0,477
74	84	3	0,036	0,036	0,395	0,328 a 0,462
75	81	6	0,074	0,074	0,366	0,300 a 0,432
76	75	2	0,027	0,027	0,356	0,290 a 0,422
77	73	6	0,082	0,082	0,327	0,263 a 0,391
78	67	4	0,060	0,060	0,307	0,244 a 0,370

Tabela 22- Resultados da análise de sobrevivência para a população masculina de Bebedouro.

(conclusão)						
Tempos de falhas (t_j)	Sobreviventes em t_j	Falhas (d_j)	Probabilidades de falhas em t_j ($\hat{q}_j$)	Taxas de falhas entre t_{j-1} e t_j	Sobrevivências acumuladas	Intervalo de confiança de 95%
79	63	4	0,064	0,063	0,288	0,226 a 0,350
80	59	4	0,068	0,068	0,268	0,208 a 0,329
81	55	6	0,109	0,109	0,239	0,181 a 0,297
82	49	3	0,061	0,061	0,224	0,167 a 0,281
83	46	4	0,087	0,087	0,205	0,150 a 0,260
84	42	2	0,048	0,048	0,195	0,141 a 0,249
85	40	4	0,100	0,100	0,176	0,123 a 0,228
86	36	6	0,167	0,167	0,146	0,098 a 0,195
87	30	4	0,133	0,133	0,127	0,081 a 0,172
88	26	7	0,269	0,269	0,093	0,053 a 0,132
89	19	4	0,211	0,211	0,073	0,037 a 0,109
90	15	6	0,400	0,400	0,044	0,016 a 0,072
91	9	1	0,111	0,111	0,039	0,012 a 0,065
92	8	1	0,125	0,125	0,034	0,009 a 0,059
93	7	1	0,143	0,143	0,029	0,006 a 0,052
95	6	2	0,333	0,167	0,020	0,001 a 0,038
96	4	2	0,500	0,500	0,010	0,000 a 0,023
97	2	1	0,500	0,500	0,005	0,000 a 0,014
107	1	1	1,000	0,100	0,000	0,000 a 0,000

Fonte: dados da análise

Tabela 23- Razões das taxas de falhas (multiplicada por 1000) da população masculina pela feminina do município de Bebedouro, SP.

(continua)			
Idades Exatas	Taxas de falhas masculinas (x1000)	Taxas de falhas femininas (x1000)	Razões das taxas de falhas masculinas pela femininas
0	0,000	0,000	-
1	60,976	0,000	-
2	60,976	0,000	-
3	60,976	0,000	-
4	60,976	0,000	-
5	60,976	0,000	-
6	60,976	0,000	-
7	60,976	0,000	-
8	60,976	0,000	-
9	60,976	0,000	-
10	60,976	0,000	-
11	60,976	0,000	-
12	0,411	0,000	-
13	0,411	0,000	-

Tabela 23- Razões das taxas de falhas (multiplicada por 1000) da população masculina pela feminina do município de Bebedouro, SP.

(continua)

Idades Exatas	Taxas de falhas masculinas (x1000)	Taxas de falhas femininas (x1000)	Razões das taxas de falhas masculinas pela femininas
14	0,411	0,000	-
15	0,411	0,000	-
16	0,411	0,000	-
17	0,411	0,000	-
18	1,642	0,000	-
19	1,642	0,278	5,906
20	1,642	0,278	5,906
21	1,642	0,278	5,906
22	1,244	1,862	0,668
23	1,244	1,862	0,668
24	1,244	1,862	0,668
25	1,244	1,862	0,668
26	2,500	1,404	1,781
27	5,050	1,404	3,597
28	5,076	1,404	3,615
29	20,408	1,404	14,536
30	20,408	1,404	14,536
31	20,408	1,130	18,060
32	20,408	5,682	3,592
33	2,604	5,682	0,458
34	10,526	2,860	3,680
35	10,526	5,747	1,832
36	2,659	5,780	0,460
37	2,659	5,814	0,457
38	2,659	5,814	0,457
39	3,565	2,924	1,219
40	3,565	5,882	0,606
41	5,405	5,882	0,919
42	5,405	5,882	0,919
43	8,197	1,972	4,157
44	5,556	1,972	2,817
45	5,586	2,976	1,877
46	16,854	2,976	5,663
47	16,854	2,976	5,663
48	2,857	2,976	0,960
49	11,494	2,976	3,862
50	17,442	3,593	4,854
51	11,834	12,195	0,970
52	5,988	18,518	0,323
53	12,048	18,518	0,651
54	18,293	3,145	5,817

Tabela 23- Razões das taxas de falhas (multiplicada por 1000) da população masculina pela feminina do município de Bebedouro, SP.

(continua)

Idades Exatas	Taxas de falhas masculinas (x1000)	Taxas de falhas femininas (x1000)	Razões das taxas de falhas masculinas pela femininas
55	24,845	6,329	3,926
56	25,478	31,847	0,800
57	26,144	6,579	3,974
58	53,691	6,622	8,108
59	53,691	20,000	2,685
60	17,730	20,408	0,869
61	36,765	20,833	1,765
62	15,267	7,092	2,153
63	31,008	14,286	2,171
64	24,000	7,246	3,312
65	24,590	7,299	3,369
66	25,210	7,353	3,429
67	34,483	22,222	1,552
68	44,643	15,151	2,947
69	56,075	15,385	3,645
70	39,604	62,500	0,634
71	82,474	58,333	1,414
72	22,472	26,549	0,846
73	34,483	9,091	3,793
74	35,714	36,697	0,973
75	74,074	57,143	1,296
76	26,667	50,505	0,528
77	82,192	31,915	2,575
78	59,702	65,934	0,905
79	63,492	58,823	1,079
80	67,797	25,000	2,712
81	109,091	38,461	2,836
82	61,224	66,667	0,918
83	86,956	71,428	1,217
84	47,619	107,692	0,442
85	100,000	103,448	0,967
86	166,667	96,154	1,733
87	133,333	42,553	3,133
88	269,231	44,444	6,058
89	210,526	162,791	1,293
90	400,000	166,667	2,400
91	111,111	200,000	0,556
92	125,000	166,667	0,750
93	142,857	100,000	1,429
94	142,857	333,333	0,429
95	166,667	333,333	0,500

Tabela 23- Razões das taxas de falhas (multiplicada por 1000) da população masculina pela feminina do município de Bebedouro, SP.

(conclusão)

Idades Exatas	Taxas de falhas masculinas (x1000)	Taxas de falhas femininas (x1000)	Razões das taxas de falhas masculinas pela femininas
96	500,000	375,000	1,333
97	500,000	400,000	1,250
98	100,000	333,333	0,300
99	100,000	333,333	0,300
100	100,000	250,000	0,400
101	100,000	250,000	0,400
102	100,000	250,000	0,400
103	100,000	333,333	0,300
104	100,000	333,333	0,300
105	100,000	0,000	-
106	100,000	0,000	-
107	100,000	0,000	-

Fonte: dados da análise