

Relação entre a expectativa de vida, período de gestação e média de descendentes em mamíferos como forma de preservação da espécie

Raphaella Tamyres Siqueira Gonçalves and Heitor Ribeiro Bernardes

Illum Escola de Ciência

6 de outubro de 2023

Resumo

Relatório apresentado à disciplina de Ciências ambientais do quarto semestre do curso de bacharelado em Ciência, Tecnologia e Inovação da Illum, escola de ciência, sob orientação do Prof. Dr. Leandro Nascimento Lemos. O objetivo é identificar a relação entre a expectativa de vida, o período de gestação e a média de descendentes mamíferos, para inferir sobre a dinâmica de preservação da espécie, evitando um quadro de extinção. Os dados foram obtidos a partir da plataforma *Animal Diversity Web - Quardvark*. A hipótese consiste em verificar se a medida que um indivíduo de determinada espécie tem um expectativa de vida alta, o número de descendentes é menor e o período de gestação é maior, e vice-versa. Desta forma, para um indivíduo com baixa expectativa de vida, a prole deve ser maior para preservar a continuidade da sua espécie. Os resultados mostraram que há uma relação moderadamente satisfatória entre os atributos que sustentam essa hipótese, excetuando as espécies que expressam o inverso do pressuposto, identificadas como *outliers*.

1 Introdução

Ao longo da história evolutiva, os seres vivos vêm sofrendo diversas mutações, para se adaptarem às condições ambientais e fatores que impactam na sobrevivência da espécie. Em decorrência disso, múltiplas estratégias de reprodução foram emergindo, mesmo quando comparando indivíduos de uma mesma classe.

Este relatório tem por objetivo verificar a relação entre atributos que estejam relacionados com a preservação das espécies, afastando-as de um possível quadro de extinção. É estabelecida a hipótese de que a expectativa de vida está relacionada com o número de decentes que um indivíduo gera, bem como o período de gestão influencia nesse processo, para manter a continuidade de indivíduos na população.

Em um estudo de correlações, o grupo de Jiaqi Wu encontrou um fator de correlação de até 0.726 entre o período de gestão e a expectativa de vida em mamíferos placentários, o que suporta a hipótese levantada neste trabalho. [2]

2 Metodologia

Para testar a hipótese, é necessário fazer um levantamento de dados que permita verificar as correspondências entre os elementos observados. Foi utilizado o banco de dados disponibilizado pela plataforma *Animal Diversity Web - Quardvark*. Os critérios de busca definidos foram categorizados em classe e condições da seguinte forma.

Classe:

- *Animal Group: Mammalia*

Condições:

- *Lifespan / Longevity: Lifespan*
- *General Behavior: Gestation period*
- *General Behavior: Number of offspring*

A partir desses critérios, foi obtido um *data frame* em formato .xls com 651 dados. Os dados foram processados, tratados e analisados utilizando o *Jupyter Notebook*, interpretador da linguagem Python versão 3.9.7.

Foi utilizada a biblioteca *Scikit-learn* durante o processo de explorar os dados, pois ela fornece ferramentas estatísticas, como o PCA (Análise de Componentes Principais), a biblioteca *Seaborn* que gera gráficos estatísticos, como a Correlação de Pearson para verificar a semelhança ou dissimilaridade entre os atributos, e o *Matplotlib* para visualização de gráficos como de dispersão, com o intuito de entender o comportamento dos dados.

Durante o processamento dos dados foi identificado um elemento faltante (NaN - *Not a Number*) e este foi removido, sendo analisadas 650 espécies. O atributo referente ao período de gestação foi obtido em minutos, sendo realizada a conversão para dias. Assim, os dados não apresentaram grande discrepância em termos de unidade, podendo ser analisados sem a necessidade de normalização dos atributos, visto que variavam em no máximo três casas decimais entre eles, o que é aceitável.

3 Resultados

A seguir são apresentados os resultados com base na correlação de Pearson, no gráfico de dispersão, na variância acumulada, no PCA, detecção de outlier e tendência dos dados.

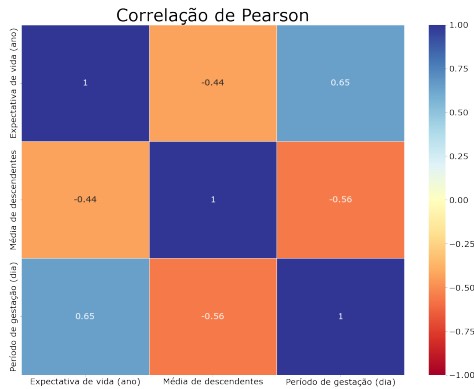


Figura 1: Correlação de Pearson

A correlação de Pearson é uma ferramenta estatística utilizada para avaliar a relação linear entre duas ou mais variáveis em um conjunto de dados. Ela se expressa numa escala entre -1 e 1, descrevendo a força e a direção da relação linear entre as variáveis, sendo que valores próximos a -1 indicam alta correspondência negativa, em que as variáveis se movem em direções opostas, enquanto valores próximos a 1 apontam para uma elevada correspondência positiva, em que as variáveis tendem a aumentar juntas. Valores próximos de zero exprimem uma baixa correlação linear entre as variáveis, ou seja, seus valores não aumentam ou diminuem juntos de maneira consistente.

A correlação de Pearson (figura 1) revela que há uma razoável relação diretamente proporcional entre as variáveis período de gestação e expectativa de vida, e uma relação moderadamente inversa entre expectativa de vida e média de descendentes e entre esta última e o período de gestação.

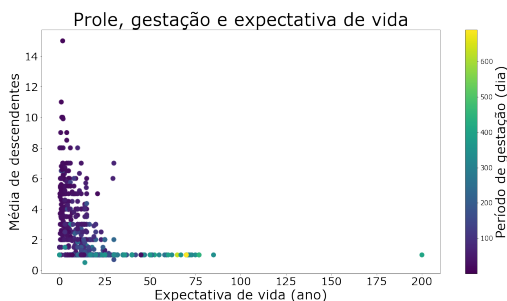


Figura 2: Gráfico de dispersão

O gráfico de dispersão (figura 2) apresenta a relação entre os três atributos analisados, sendo a expectativa de vida e a média de descendentes em relação ao período de gestação. Observa-se que há

uma maior concentração dos dados entre aproximadamente 0-20 anos, com uma média de descendentes entre 1-7 e um período de gestação em sua maioria até cerca de 200 dias.

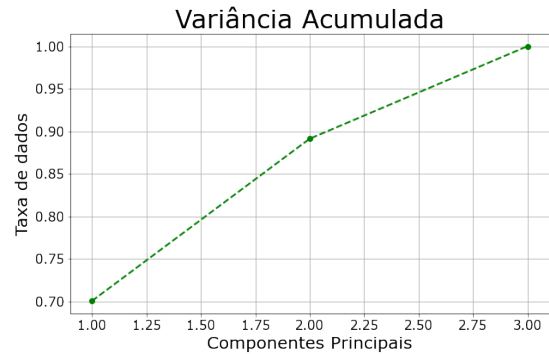


Figura 3: Gráfico de variância

O gráfico de variância acumulada é uma métrica usada para quantificar quanto da variabilidade total nos dados é explicada por um determinado número de componentes principais, sendo essencial para ajudar a determinar quantos desses componentes principais devem ser retidos para representar eficazmente os dados originais.

Analisando o gráfico (figura 3), nota-se que as três componentes principais são significativas para a variação dos dados, pois a reta não apresenta uma súbita mudança de direção ou declínio.

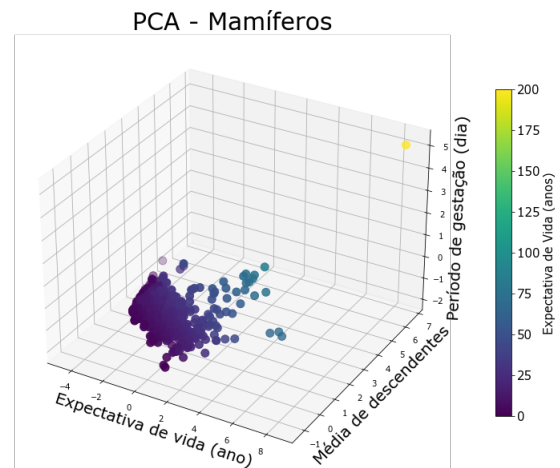


Figura 4: Análise de componentes principais - PCA

O PCA (*Principal Component Analysis*) é uma técnica estatística utilizada para a redução de dimensionalidade e a análise de dados multivariados. O objetivo principal do PCA é transformar um conjunto de variáveis correlacionadas em um conjunto menor de variáveis não correlacionadas, chamadas de componentes principais, de maneira que a maior parte da variabilidade original dos dados seja preservada.

Como observado na figura 4, a distribuição dos pontos evidencia a formação de um agrupamento

das três componentes principais em uma determinada região, indicando que existe uma característica semelhante entre elas e as três podem ser retidas nesse estudo.

Assim como no PCA, a figura 5 confirma a existência de alguns dados considerados anômalos, ou seja, que apresentam baixa densidade local em relação aos vizinhos, contrariando o comportamento esperado, com base nos outros pontos observados.

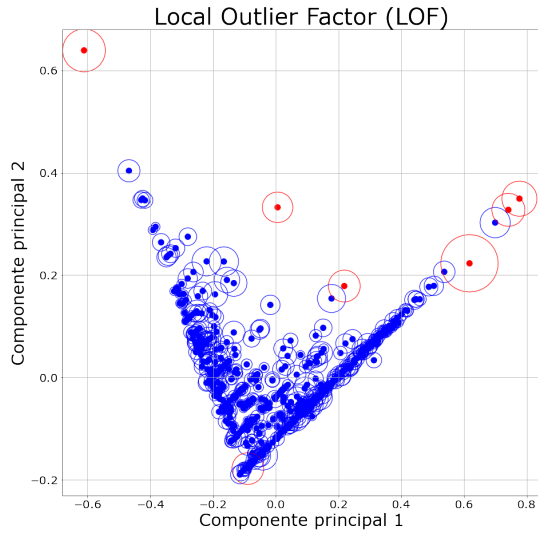


Figura 5: Local Outlier Factor (LOF)

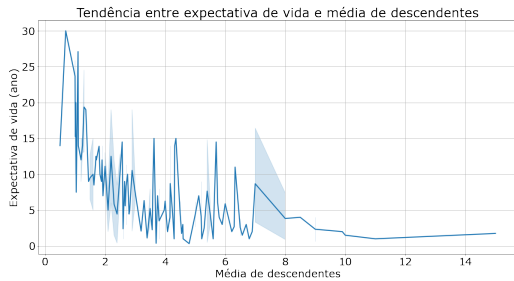


Figura 6: Tendência entre expectativa de vida e a média de descendentes

No gráfico de tendência entre a expectativa de vida e a média de descendentes (figura 6), observa-se que há um declínio entre os pontos de flutuação, demonstrando que a medida em que há uma expectativa de vida baixa, é preciso que a média de descendentes aumente, favorecendo a continuidade da espécie.

No gráfico de tendência entre expectativa de vida e período de gestação (figura 7), observa-se que enquanto o período de gestação tende a aumentar, os indivíduos apresentam maior expectativa de vida.

Em ambos os gráficos, constata-se que os dados não são totalmente lineares, devido aos *outliers*, mas isso não interfere significativamente na compreensão de padrões de redução e aumento da expectativa de vida, conforme as variáveis analisadas.

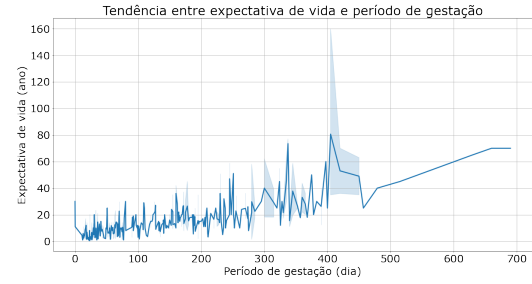


Figura 7: Tendência entre expectativa de vida e período de gestação

4 Discussão

Ao analisar o gráfico de correlação de dispersão, percebe-se que os indivíduos com expectativa de vida maior que 40 anos tendem a ter, em média, 1 descendente em sua vida; enquanto é mais provável encontrar indivíduos com expectativa de vida menor que 25 anos que possuem prole esperada acima de 6 indivíduos.

Ao analisar os resultados, verifica-se uma correlação de -0.44 , o que é um valor moderado e indica que são inversamente relacionados. Além disso, é interessante observar que há uma correlação de 0.65 entre expectativa de vida e período gestacional, o que está congruente com os valores encontrados na literatura. Não obstante, foi encontrada ainda uma relação inversa moderada entre a média de descendentes e o período gestacional, com valor de -0.56 , o que também pode ser observado no mapa de cores do gráfico de dispersão mencionado anteriormente.

Diante do exposto, é notável que de fato pode-se traçar correlações entre o período gestacional, a expectativa de vida e quantos descendentes um mamífero tem em sua vida, sustentando assim as hipóteses levantadas no começo deste trabalho. Ainda assim, deve-se ressaltar que estes não são os únicos atributos que interagem entre si, bem como não são as únicas condições relacionadas com a preservação de uma espécie. Fatores como massa cerebral, tipo de dieta, tamanho de grupo, densidade populacional, taxa de migração, dentre diversos outros, também podem estar conectados com as observações aqui feitas; contudo os três fatores deste trabalho foram selecionados utilizando um critério de espaço amostral conforme a disponibilidade de dados no *Quaardvark*.

Referências

- [1] JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; et al. **An introduction to statistical learning: with applications in Python**. Cham, Switzerland: Springer, 2023.
- [2] WU, Jiaqi; YONEZAWA, Takahiro; KISHINO, Hirohisa. **Evolution of Reproductive Life History in Mammals and the Associated Change of Functional Constraints**. Genes, v. 12, n. 5, p. 740, 2021.