Predicting a Lemur's Lifespan in Wild and Captivity*

Khushaal Nandwani

November 27, 2024

This study aims to predict the lifespan of lemurs in wild and captive by analyzing key factors such as sex, species, genus, and birth month. Utilizing data from the Duke Lemur Center, I divided the dataset into wild-born and captive-born lemurs to assess differences in longevity. I employed Generalized Linear Models (GLMs) with a Gaussian distribution to model lifespan based on the selected variables. My findings indicate that lemurs in captivity have a significantly higher average lifespan compared to their wild counterparts, likely due to controlled environments and access to medical care. In wild populations, sex was a significant predictor of lifespan, with males living longer than females, whereas in captivity, sex did not have a substantial impact. Species differences were prominent in both settings, suggesting inherent biological factors affecting longevity. Birth month showed inconsistent effects on lifespan across both groups. These results highlight the influence of environmental conditions on lemur longevity and underscore the importance of species-specific conservation strategies.

Table of contents

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

1 Introduction

Lemurs are the world's most endangered group of mammals (Duke Lemur Center (2024)). Understanding the factors that influence their lifespan is crucial for guiding effective conservation efforts. The Duke Lemur Center (DLC), established in 1966, has played a pivotal role in the study and preservation of lemurs, maintaining the largest population of these species outside their native Madagascar. Different factors can affect the longevity of lemurs in wild and captive, longevity is important because the longer lemurs live, the more likely they are to reproduce and contribute to the survival of their species.

This study uses a dataset from the DLC to analyze if and how sex, species, genus and month a lemur is born in can affect its lifespan in the wild and captivity. This study can help us guide what kinds of lemur are we failing to conserve in captivity and how they perform in wild. It will also reinforce where our efforts have been successful.

Using Generalized Linear Models (GLMs) with Gaussian distribution, factors including sex, species, genus, and birth month were evaluated to predict lemur longevity. The dataset was divided into wild-born and captive-born lemurs to understand the differences in lifespan between the two environments and separate models were built for each group.

The findings reveal that lemurs in captivity exhibit significantly longer lifespans. In wild populations, males tend to live longer than females, while species differences are prominent in both environments, emphasizing the need for species-specific conservation strategies. Birth month was found to have inconsistent effects across both groups, suggesting complex interactions with environmental conditions.

The remainder of this paper is structured as follows: Section 2 details the data collection and processing methods, while Section 3 outlines the modeling approach. Section 4 presents the results, highlighting key differences between wild and captive lemurs. The discussion in **?@sec-sec-discussion** contextualizes these findings, addressing limitations and proposing future directions for research. Finally, the **?@sec-appendixs** includes additional information on data cleaning, the analysis dataset, and the model dataset.

2 Data

The data was taken from Cookson (2020), who acquired it from Duke Lemur Center more about which, can be found in Section 2.1. We used R Core Team (2023)

It is important to know about Lemurs and what affects their lifespan because they are the most endangered mammals on the planet. The Duke Lemur Center (DLC) is a global leader in the research, care, and conservation of lemurs. The DLC hosts the most diverse population of lemurs outside their native habitat in Madagascar.

To do so, I chose the following variables from the dataset, which I believe are the most important factors that can affect the lifespan of a lemur:

- `Sex`: The sex of the lemur on birth. It is a cateogrical variable that can be `M` or `F`.
- `Species`: The species of the lemur. It is also a categorical variable and can take one of the following values: `GG`, `COL`, `UL`, `RUF`, `MOH`, `MAC`, `CAT`, `FUL`, `ALB`, `AR`, `VV`, `COQ`, `MED`, `MUR`, `COU`, `TAR`, `PYG`, `ZAZ`, `MON`, `RUB`, `COR`, `SAN`, `FLA`, `MAD`, `POT`. The specific or common names of these species can be found in Appendix A.
- `Genus`: The genus of the lemur. It is a categorical variable and can take one of the following values: `O`, `E`, `G`, `H`, `L`, `V`, `P`, `C`, `M`, `N`, `D`.
- `month_born`: The month in which the lemur was born. It is a categorical variable and can take one of the following values: `1`, `2`, `3`, `4`, `5`, `6`, `7`, `8`, `9`, `10`, `11`, `12`, where `1` corresponds to January, `2` to February, and so on.

Finally, we have the target variable `Age` which is the lifespan of the lemur in years. It is a continuous variable.

The following plots show the distribution of the lemurs in the dataset based on the variables `Sex`, `Species`, `Genus`, and `month_born` for both lemurs in captivity and in the wild.

For `Sex`, we observe that in the wild, the distribution of male and female lemurs is nearly even, with only a slight skew. While in captivity, there are significantly more female lemurs than males. The larger count of female lemurs in captivity could result from conservation breeding programs prioritizing females to ensure the species' survival.

Looking at the distribution of `Species`. We observe that in the wild, the distribution is relatively balanced, with a few exceptions of `MAD`, `CAT` and `MED` which appear to be low in number. In captivity, certain species (e.g., MUR, CAT, MED) are more represented, while others are rare or absent. This can be due to captive environments may focus on saving the most endangered species, as we see the ones least in wild like `CAT` and `MED` are more in captivity.

With the exception of genus `E`, genus representation is balanced in both wild and captive lemurs. `E` might be high in both cases, as it could be a common genus or have a higher conservation priority.

For birth month, births in the wild are concentrated in a few months, suggesting a breeding season. While in captivity, births are more evenly spread across the year, with some peaks in specific months. Seasonal breeding in the wild is influenced by environmental factors like food availability and climate. Captivity can disrupt these natural cycles, as controlled environments and year-round resources allow for more frequent and less seasonal reproduction.

For `Age`, we observe that in the wild, the age distribution of lemurs shows a higher frequency of younger lemurs, with relatively few individuals reaching older ages. In captivity, the age distribution extends further, with a significant number of lemurs living to older ages. This difference can be attributed to the controlled environments in captivity, which provide consistent food, medical care, and protection from predators, thus increasing longevity. In contrast, the challenges of the wild, such as predation, disease, and fluctuating resources, contribute to shorter lifespans. With there being one exception, we observe signigicant number of captive lemurs below age 1, this can be due to the extensive breeding of captivity or could be a sign of data quality issues.

We will also look at the distribution of the target variable `Age` in the dataset. The following plot shows the distribution of the ages of the lemurs in the dataset for both lemurs in captivity and in the wild.

2.1 Data Collection

The data in these sources was acquired and processed by staff at the Duke Lemur Center (DLC).

2.1.1 Data Acquisition

As Zehr et al. (2014) points out, DLC staff collected data about the lemurs according to standard operating procedures and USDA, AZA, and IACUC guidelines. They recorded information about births, deaths, weights, enclosure moves, behaviors, and other significant events

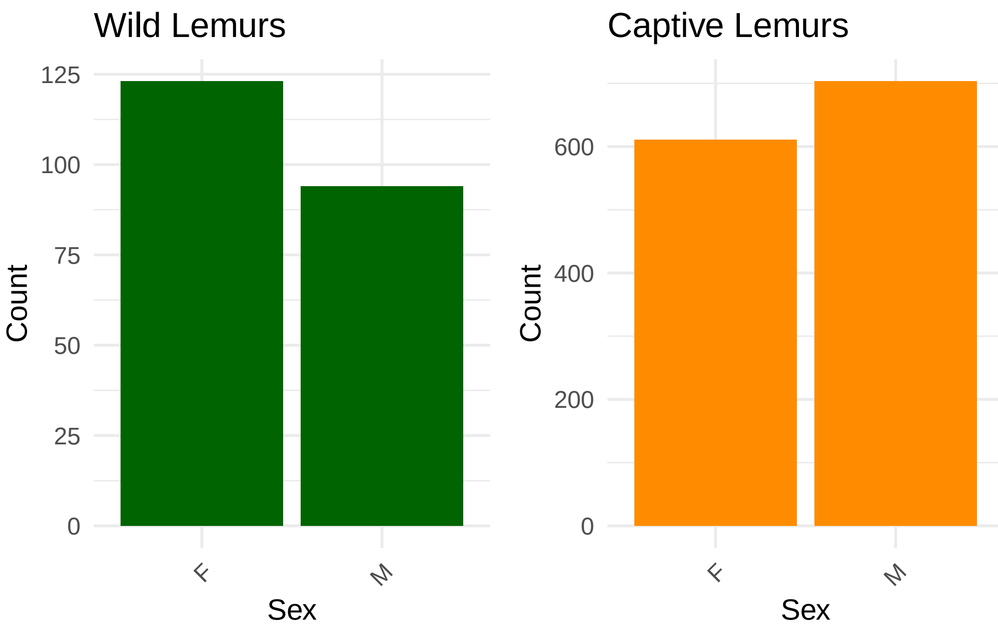

Figure 1: Distribution of Lemurs by Sex

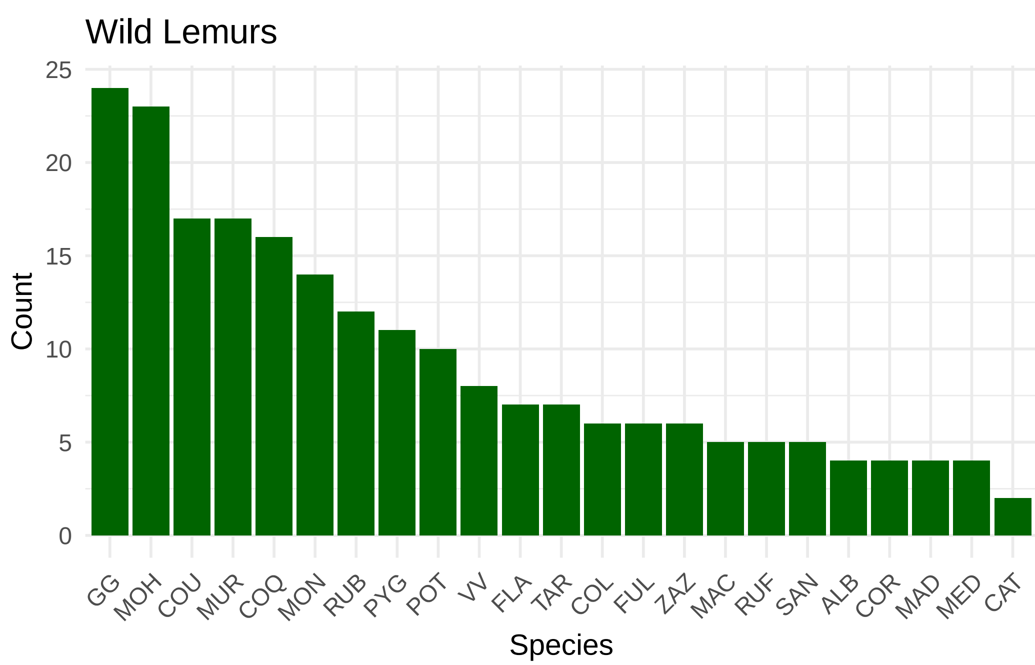

Figure 2

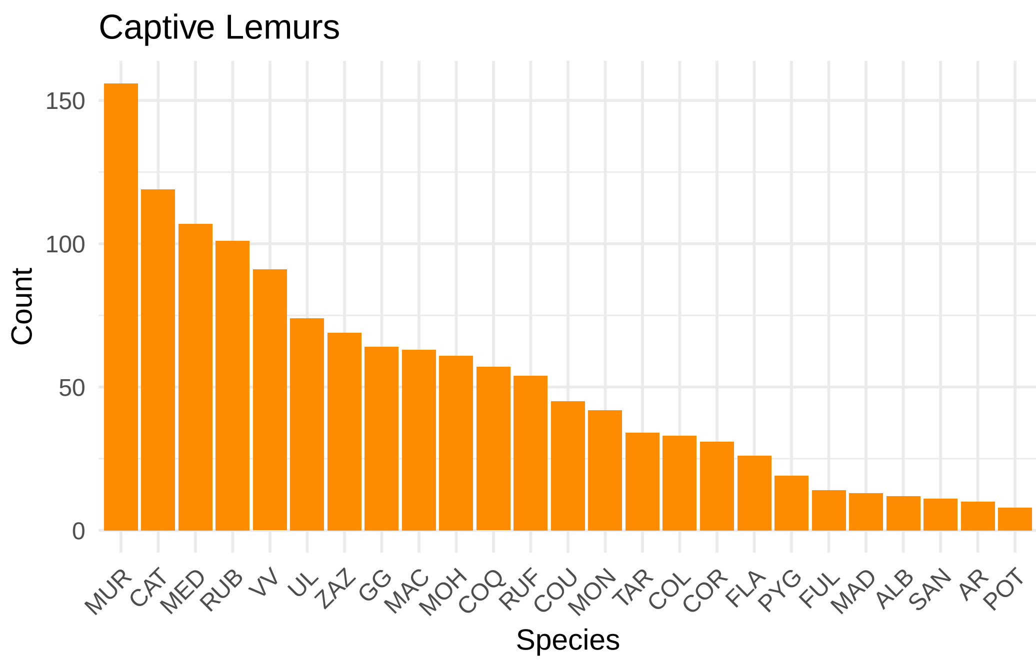

Figure 3

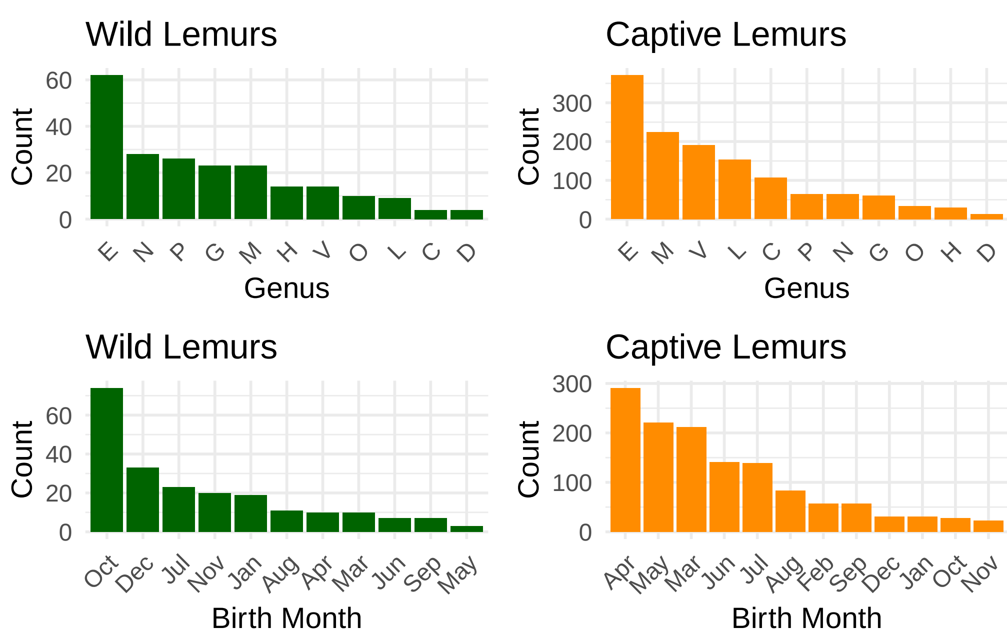

Figure 4

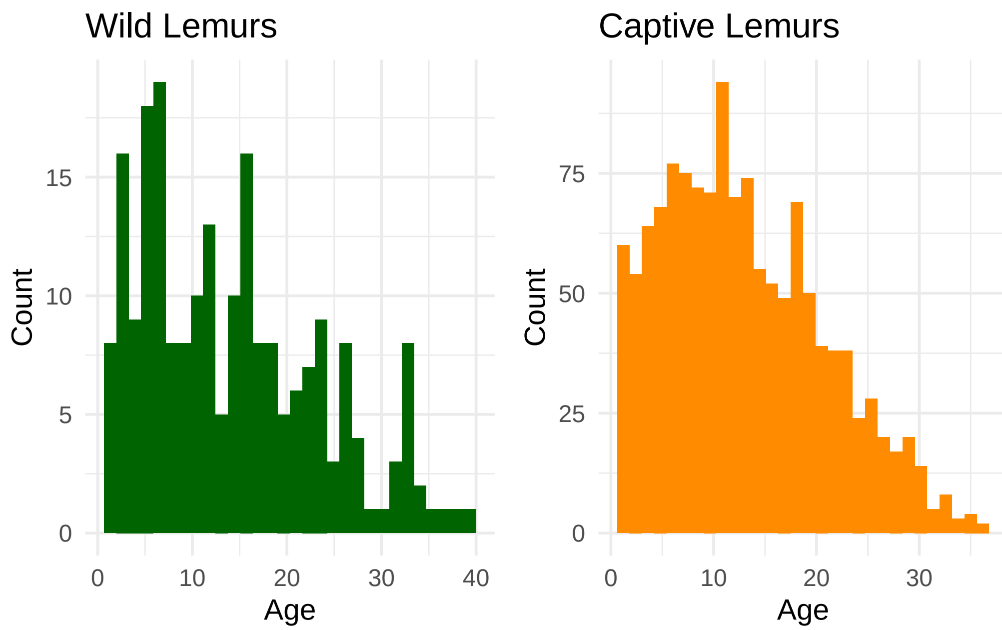

Figure 5: Distribution of ages of lemurs in the dataset for both lemurs in captivity and in the wild.

on a daily basis. Originally, this data was stored in handwritten and typed paper formats. Later, it was computerized.

In the mid-1990s, the DLC started using two databases: the Animal Record Keeping System (ARKS) and MedARKS. These databases allowed the DLC to share information with other organizations through the International Species Information System (ISIS). The DLC is currently transitioning to using the Zoological Information Management System (ZIMS). Data not stored in these databases has been stored in spreadsheets, and the DLC is working on transferring data from older records into these databases.

2.1.2 Data Processing

The DLC used SAS software to build a database for the lemur data. Data from various sources was imported into SAS Enterprise Guide, including ARKS, MedARKS, ZIMS, and spreadsheets. They wrote programs in SAS to extract, match, and join data, calculate new variables, and format the output. They also used tools within SAS Enterprise Guide Projects for calculations and formatting. The DLC uses a unique ID to match data for individual animals, and the taxonomic name for species-related variables. The data was validated by identifying and locating missing data, standardizing codes and text, investigating outliers, and comparing known values to the database output.Data that could not be verified was excluded from the published dataset.

The DLC created two data files from the database: **the DLC Animal List**, which contains single-copy variables for each animal in the colony's history, and **the DLC Weight File**, which contains all weight measurements for each animal. We used the DLC Animal List for this analysis, which was first cleaned by Cookson (2020) and then we clean it as per our needs.

The data in these sources was updated on February 8, 2019. The DLC plans to update the data on a yearly basis.

3 Model

I used Generalized Linear Model (GLM) framework with a Gaussian family distribution to determine the expected lifespan of a lemur based on the selected variables. GLM is a statistical model that generalizes linear regression to include non-normal distributions. It is used when the dependent variable is not normally distributed or when the relationship between the dependent and independent variables is not linear. The Gaussian family distribution is used when the dependent variable is continuous and normally distributed, which was in my case.

My model will be based on `Sex`, `Species`, `Genus`, and `month_born` as the independent variables and `Age` as the dependent variable.

Mathematically, the model can be represented as:

$$\bar{a} = \beta_0 + \beta_1 \times \texttt{Sex} + \beta_2 \times \texttt{Species} + \beta_3 \times \texttt{Genus} + \beta_4 \times \texttt{month\_born} + \epsilon \tag{1}$$

$$\begin{aligned}
\bar{a}_{\text{wild}} &\sim \text{Normal}(16, 4)\\
\bar{a}_{\text{captive}} &\sim \text{Normal}(26, 4)\\
\beta_0 &\sim \text{Normal}(0, 2.5)\\
\beta_1 &\sim \text{Normal}(0, 2.5)\\
\beta_2 &\sim \text{Normal}(0, 2.5)\\
\beta_3 &\sim \text{Normal}(0, 2.5)\\
\beta_4 &\sim \text{Normal}(0, 2.5)
\end{aligned}$$

where,

- $\bar{a}_{\text{wild}}$ is the expected lifespan of a lemur in the wild.
- $\bar{a}_{\text{captive}}$ is the expected lifespan of a lemur in captivity.
- β_0 is the intercept term. It represents the expected lifespan of a lemur when all other variables are zero.
- β_1 is the coefficient for `Sex`. It represents the change in the expected lifespan of a lemur for a one-unit change in `Sex`.

- β_2 is the coefficient for `Species`. It represents the change in the expected lifespan of a lemur for a one-unit change in `Species`.
- β_3 is the coefficient for `Genus`. It represents the change in the expected lifespan of a lemur for a one-unit change in `Genus`.
- β_4 is the coefficient for `month_born`. It represents the change in the expected lifespan of a lemur for a one-unit change in `month_born`.
- ϵ is the error term.

We chose a Gaussian family distribution for the model because the dependent variable `Age` is continuous and normally distributed. From Cape May County, NJ (2024) we found that the expected lifespan of a lemur in wild is upto 18 years, and in captivity about 30 years. So, we set the prior for the expected lifespan of a lemur in the wild to be $\text{Normal}(16, 4)$ and in captivity to be $\text{Normal}(26, 4)$. We set the priors for the coefficients to be $\text{Normal}(0, 2.5)$, because we do not have any prior information about the effect of the independent variables on the dependent variable.

Using moderately broad priors can help regularize the model, reducing the risk of overfitting and make us have more stable estimates, especially in scenarios with limited data.

(**app_model-dataset?**) covers coefficients, standard errors, convergence checks and other model diagnostics for our model.

4 Results

Figure 6 and Figure 7 shows the coefficients and their 95% credible intervals for the predictor variables in the respective wild and captive generalized linear regression models. A 95% credible interval indicates a 95% chance that the true parameter value falls within the interval, based on the observed data and model assumptions. Each point represents the estimated coefficient for a predictor, and the horizontal lines show the credible interval around it. Coefficients to the right of zero suggest a positive relationship with the outcome variable, meaning that as the predictor increases, the outcome also increases. Coefficients to the left of zero suggest a negative relationship, meaning that as the predictor increases, the outcome decreases.

4.1 Wild Lemurs

Few species in the wild have significant effect on the age, these include `VV`, `SAN`, `RUF`, `RUB`, `MAD` and `CAT`. For sex, as expected in wild, males have a higher lifespan than females. We had to jump over p-value of 0.07 to get the significant results, as we had limited data in case of wild.

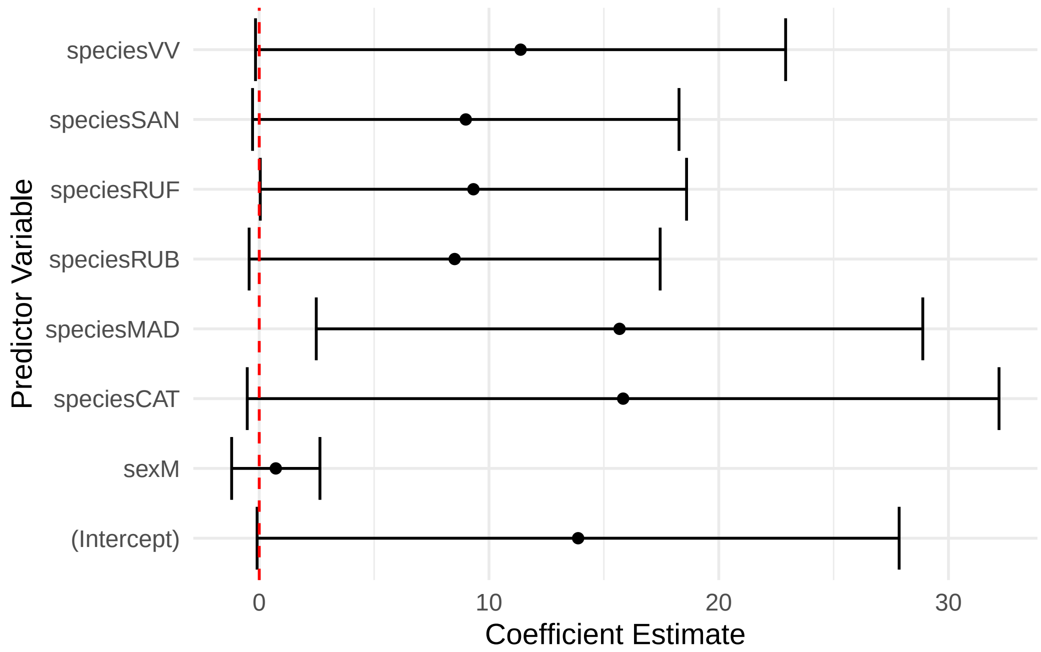

Figure 6: Coefficients of the wild model

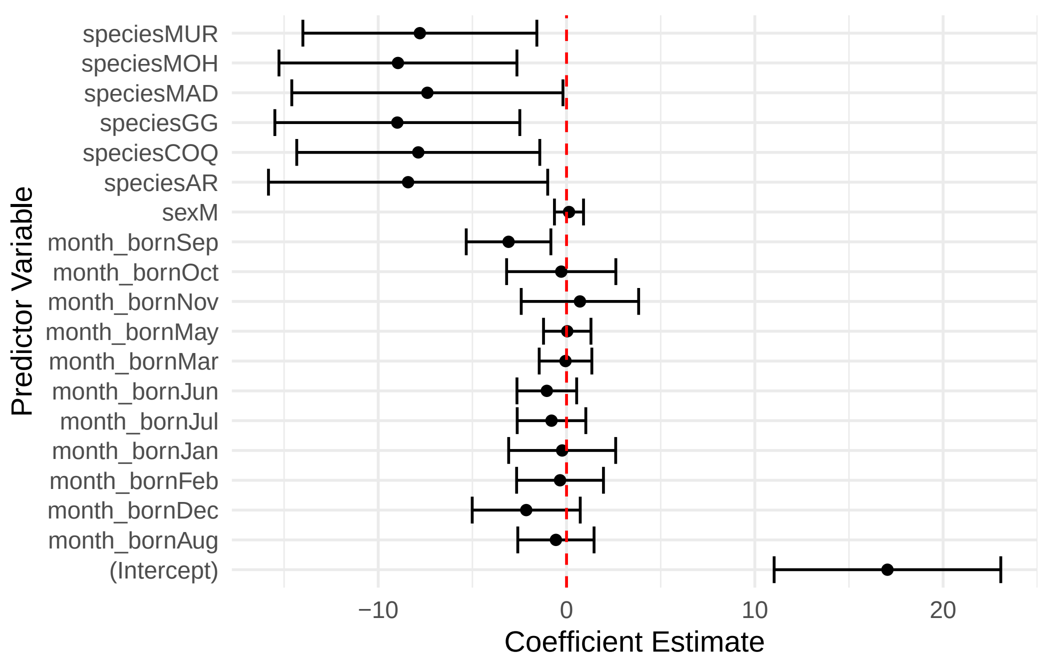

Figure 7: Coefficients of the captive model

4.2 Captive Lemurs

Few species have a negative effect on age, these include `MUR`, `MOH`, `MAD`, `GG`, `COQ` and `AR`. In case of captivity, sex is not a major factor in age, as its confidence interval is nearly centered at 0. Overall, species differences are prominent, while birth month effects appear less consistent, mostly centered around 0. Except December and September seem to be negatively affecting the age of lemurs in captivity, while November positively.

4.3 Effect of Sex on Age

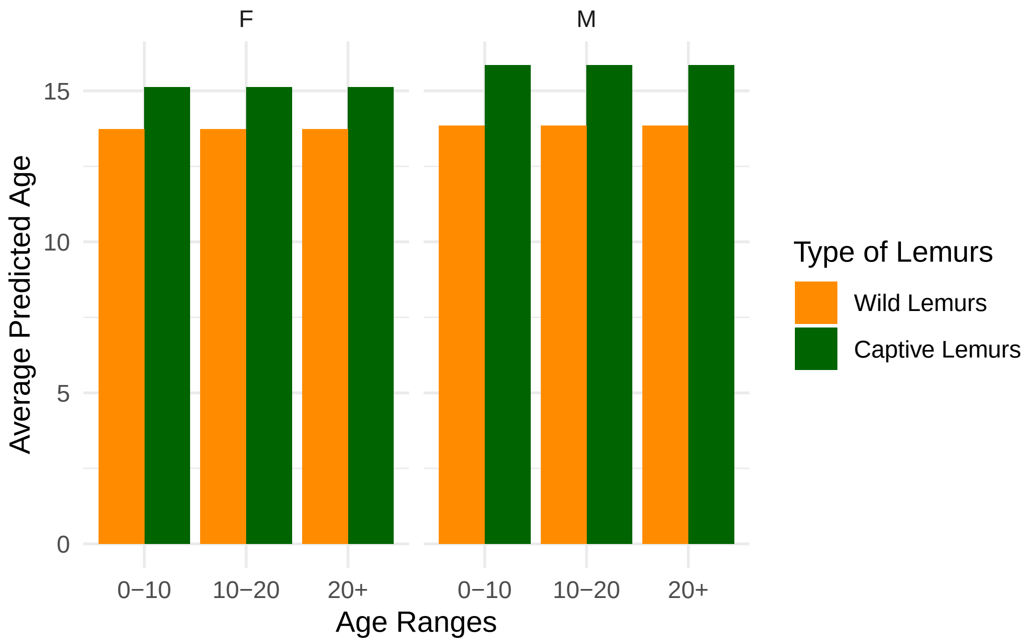

Figure 8: Comparison of Average Predicted Ages of Lemurs by Age Range, Sex, and Prediction Type (Wild vs. Captive)

From the Figure 8, lemurs in captive have a higher average predicted age than those in the wild across all age ranges. And in the captive, we see males living longer on average than females, while in the wild, the difference is less pronounced. This can be due to the controlled environment in captivity, which provides consistent food, medical care, and protection from predators, thus increasing longevity. In the wild, the challenges of predation, disease, and fluctuating resources contribute to shorter lifespans.

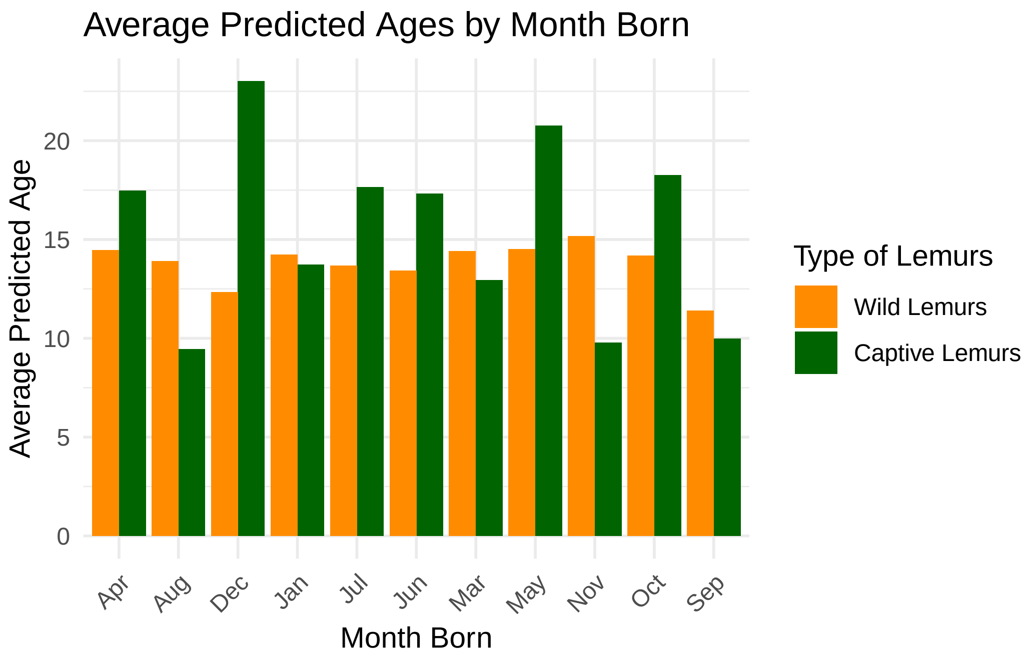

Figure 9: Comparison of Average Predicted Ages of Lemurs by Age Range, Birth Month, and Prediction Type (Wild vs. Captive)

4.4 Effect of Month Born on Age

For months in Figure 9, we see captive lemurs outlive wild lemurs most of the months except August, January, March, Novemeber and September. It is worth exploring the reasons behind these differences by Duke Lemur Center, as they can provide insights into the impact of seasonal factors on lemur lifespan.

4.5 Effect of Species on Age

Once again in species as seen in Figure 10, we see captive lemurs living longer than wild ones in most of them, except `ALB`, `COR`, `MAC` and `MON`. Why these species are living longer in wild, can be a good area of research for Duke Lemur Center.

5 Discussion

In this study, we sought to predict and analyze the lifespan of lemurs in wild and captive environments by examining factors of sex, species, genus, and birth month. Utilizing data from the Duke Lemur Center, we divided the dataset into wild-born and captive-born lemurs to understand how environmental conditions impact longevity. Generalized Linear Models

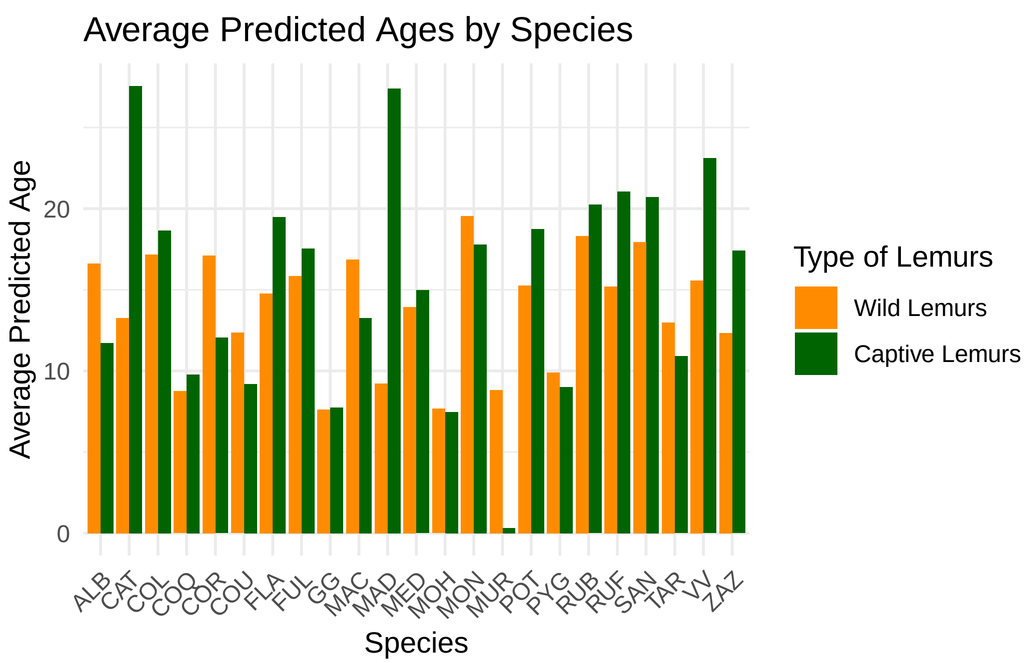

Figure 10: Comparison of Average Predicted Ages of Lemurs by Age Range, Species, and Prediction Type (Wild vs. Captive)

(GLMs) with a Gaussian distribution were employed to understand the relationships between these variables and lifespan.

5.1 Key Findings

We learn from this study that captivity plays a major role in the lifespan of lemurs. Captive lemurs, benefit from controlled environments, consistent food supplies, and medical care, live considerably longer on average than their wild counterparts similar to as the study by Tidière et al. (2016) showed. This underscores the potential of captive breeding programs to contribute positively to species survival and highlights the role of human intervention in extending the lifespans of endangered species.

Another important finding is the variability in lifespan based on species and sex. In the wild, male lemurs tend to live longer than females, possibly due to biological or behavioral factors. However, this sex-based difference diminishes in captivity, suggesting that controlled environments mitigate sex-specific survival challenges. Additionally, species differences in longevity were significant in both settings, highlighting the influence of inherent biological factors and emphasizing the need for species-specific conservation strategies.

5.2 Weaknesses and Limitations

Despite these findings, this study has several limitations. The dataset used was relatively small for wild-born lemurs, potentially affecting the reliability of the statistical models. For instance, the limited sample size may have led to the inclusion of predictors with p-values slightly above the conventional threshold, which could inflate Type I error rates. Furthermore, the data on wild-born lemurs lacked detailed information on maternal and paternal species and litter size, which are potentially critical factors in understanding lifespan dynamics, and we could not include them in our analysis.

Another limitation lies in the potential bias introduced by data collection processes. The reliance on historical records and the transition between database systems may have resulted in inconsistencies or inaccuracies. Additionally, the study only considered certain categorical variables, excluding other possible factors such as habitat quality, diet variability, or exposure to diseases, which could provide a more holistic understanding of lemur longevity.

5.3 Future Directions

This research opens the door to several avenues for further investigation. Future studies should aim to collect larger and more representative datasets for wild-born lemurs to improve the robustness of the models. Additionally, incorporating environmental variables such as habitat degradation, climate patterns, and predation risks could provide better understanding of the challenges faced by wild populations.

Furthermore, understanding why certain species thrive better in the wild (e.g., ALB, COR, MAC, and MON) compared to captivity requires focused investigation. This could involve studying behavioral adaptations, ecological niches, and species-specific physiological traits. Expanding the analysis to include genetic data may also uncover underlying biological factors that contribute to these differences.

Finally, longitudinal studies examining the long-term impacts of captivity on lemur behavior, reproduction, and overall health would be invaluable. Integrating interdisciplinary approaches, combining ecological, genetic, and conservation sciences, could help develop more effective strategies for preserving these critically endangered species.

In conclusion, while this study provides meaningful insights into the factors influencing lemur lifespan, it also highlights the complexities of their survival in diverse environments. Addressing the limitations and expanding the scope of research will not only deepen our understanding but also support more informed conservation efforts for lemurs and other endangered species.

6 Appendix

A Data Cleaning

I took the following steps to clean the raw data and prepare it for analysis:

1. Loading the Data: The raw data is loaded from the file data/raw_data/animals.csv.
2. Column Selection: The dataset is reduced to relevant columns: `animal_id`, `taxonomic_code`, `sex`, `birth_date`, `death_date`, `birth_type`, `litter_size`, `mother_species`, and `father_species`.
3. Filtering Rows: Rows where death_date is missing are removed to ensure the data includes only animals with complete lifecycle information.
4. Extracting Genus and Species: The taxonomic_code column is split into two new columns: `genus` (first letter) and `species` (remaining characters). The original taxonomic_code column is removed as it is no longer needed.
5. Age Calculation: A new column, `age`, is calculated based on the difference between `death_date` and `birth_date`, measured in years.Lemurs with an age of less than 1 year are removed to focus on mature animals and ignore infant mortality. This was especially common for captive-born animals because excess breeding can lead to high infant mortality rates.
6. Birth Month Extraction: A new column, month_born, is created to indicate the birth month extracted from the birth_date. We needed this information to analyze seasonal effects on lifespan.
7. Splitting Data by Birth Type: The data is split into two subsets based on birth_type: one for animals born in captivity (`data_captive`) and another for those born in the wild (`data_wild`).
8. Column Removal for Wild Data: Since information on mother_species, father_species, and litter_size is unavailable for wild-born animals, these columns are removed from data_wild.
9. Handling Missing Values: Both datasets, data_wild and data_captive, are cleaned by dropping rows with missing values in key columns: sex, species, age, month_born, and genus.
10. Final Output: The cleaned datasets are saved as CSV files: Wild-born animals: `data/analysis_data/wild.csv` Captive-born animals: `data/analysis_data/captive.csv`

This systematic cleaning ensures the dataset is consistent, reliable, and ready for further analysis.

B Analysis Dataset

Table 1: Analysis dataset for Wild Lemurs

animal_id	sex	genus	species	age	month_born
0002	M	O	GG	3.93	Jan
0007	M	O	GG	7.55	Jan
0030	F	E	COL	5.55	Oct
0046	M	H	GG	1.50	Dec
0202	M	O	GG	3.93	Jan
0251	F	O	GG	2.91	Aug

Table 2: Analysis dataset for Captive Lemurs

animal_id	sex	birth_type	litter_size	genus	species	mother_father_same	age	month_born
4	M	captive	1	O	GG	1	5.94	Apr
5	M	captive	1	O	GG	1	15.47	Aug
6	F	captive	1	O	GG	1	13.59	Mar
8	F	captive	1	O	GG	1	5.38	Jul
9	M	captive	1	O	GG	1	10.38	Sep
10	M	captive	1	O	GG	1	13.46	May

Table 1 provides a glimpse of the dataset used for analysis for wild lemurs, and Table 2 provides a glimpse of the dataset used for analysis for captive lemurs.

C Model Dataset

D Species of Lemurs

The following are the species of lemurs that are present in the dataset:

Index	Code	Latin Name	Common Name
1	MED	Cheirogaleus medius	Fat-tailed dwarf lemur
2	MAD	Daubentonia madagascariensis	Aye-aye
3	ALB	Eulemur albifrons	White-fronted brown lemur
4	COL	Eulemur collaris	Collared brown lemur

Index	Code	Latin Name	Common Name
5	COR	Eulemur coronatus	Crowned lemur
6	FLA	Eulemur flavifrons	Blue-eyed black lemur
7	FUL	Eulemur fulvus	Common brown lemur
8	MAC	Eulemur macaco	Black lemur
9	MON	Eulemur mongoz	Mongoose lemur
10	RUB	Eulemur rubriventer	Red-bellied lemur
11	RUF	Eulemur rufus	Red-fronted brown lemur
12	SAN	Eulemur sanfordi	Sanford's brown lemur
13	MOH	Galago moholi	Mohol bushbaby
14	GG	Hapalemur griseus griseus	Eastern lesser bamboo lemur
15	CAT	Lemur catta	Ring-tailed lemur
16	TAR	Loris tardigradus	Slender loris
17	MUR	Mircocebus murinus	Gray mouse lemur
18	COQ	Mirza coquereli	Northern giant mouse lemur
19	COU	Nycticebus coucang	Slow loris
20	PYG	Nycticebus pygmaeus	Pygmy slow loris
21	COQ	Propithecus coquereli	Coquerel's sifaka
22	POT	Perodicticus potto	Potto
23	VAR	Varecia	Varecia hybrid
24	RUB	Varecia rubra	Red ruffed lemur
25	VV	Varecia variegata variegata	Black-and-white ruffed lemur

Cape May County, NJ. 2024. "Ring-Tailed Lemur." https://capemaycountynj.gov/1112/Ring-Tailed-Lemur#:~:text=Captive%2FWild%20Lifespan%3A%20up%20to,May%20being%20prime%20breeding%20time.

Cookson, T. Alexander. 2020. "Duke Lemur Center Dataset." https://github.com/tacookson/data/tree/master/duke-lemur-center.

Duke Lemur Center. 2024. "Duke Lemur Center." https://lemur.duke.edu.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Tidière, Morgane, Jean-Michel Gaillard, Vérane Berger, Dennis W. H. Müller, Laurie Bingaman Lackey, Olivier Gimenez, Marcus Clauss, and Jean-François Lemaître. 2016. "Comparative Analyses of Longevity and Senescence Reveal Variable Survival Benefits of Living in Zoos Across Mammals." *Scientific Reports* 6 (1): 36361. https://doi.org/10.1038/srep36361.

Zehr, Steven M., Ronald G. Roach, Danielle Haring, Julia Taylor, Fred H. Cameron, and Anne D. Yoder. 2014. "Life History Profiles for 27 Strepsirrhine Primate Taxa Generated Using Captive Data from the Duke Lemur Center." *Scientific Data* 1 (July): 140019. https://doi.org/10.1038/sdata.2014.19.