Decoding the Wild: A Statistical Exploration of Animal Traits

Assignment Title: Analysis of Animal Characteristics and Lifespan Trends
Author: Kashmira R.K.J.P.
Index : 22000862
Date: 13 Feb 2025

1. Introduction

Dataset Overview

The dataset comprises 205 animal species, capturing 16 variables that span physical, ecological, and conservation-related attributes. Key variables include:

- **Morphological Traits:** Height (cm), weight (kg), color, and average/top speeds.
- **Ecological Context:** Habitat type, predators, geographic distribution, and social structure.
- **Biological Metrics:** Lifespan (years), gestation period, and offspring per birth.
- **Conservation Status:** Categories like "Least Concern," "Vulnerable," and "Endangered."

Motivation and Objectives

Understanding animal traits and their interrelationships provides insights into evolutionary biology, ecological roles, and conservation priorities. This analysis aims to:

1. **Preprocess Data:** Clean and transform variables for analytical readiness.
2. **Explore Relationships:** Investigate how physical traits (e.g., height, weight) correlate with ecological factors (e.g., diet, habitat).
3. **Compare Lifespans:** Assess differences in lifespan across dietary categories.
4. **Statistical Validation:** Use hypothesis testing to validate observed trends.

2. Methodology

Data Preprocessing

1. **Handling Missing Values:**
    - Initial checks revealed no missing values (sum(is.na(animal_data))), allowing retention of all 205 observations.
    - *Justification:* Complete data ensured robust statistical analysis without imputation biases.
2. **Converting Ranges to Numerical Values:**
    - Columns like Height..cm. (e.g., "105-130") were split into lower/upper bounds, and midpoints were calculated.
    - Example: The range "105-130" became 117.5.
    - *Justification:* Midpoints simplify analysis while approximating central tendency, though variability within ranges is lost.
3. **Data Type Standardization:**
    - Columns initially stored as characters (e.g., Lifespan..years.) were converted to numeric formats for statistical operations.

Exploratory Data Analysis (EDA)

1. **Descriptive Statistics:**
    - Central tendency (mean, median) and dispersion (variance, standard deviation) were calculated for continuous variables.
2. **Visualizations:**
    - **Histogram:** Lifespan distribution to identify common age ranges.
    - **Scatter Plot:** Height vs. weight to explore size correlations.
    - **Box Plot:** Lifespan variation across diets (Carnivore, Herbivore, Insectivore).
3. **Correlation Analysis:**
    - Examined relationships between variables like height, weight, and speed using correlation matrices.

Hypothesis Testing

- **Independent T-test:** Designed to compare mean lifespans of carnivores and herbivores.
- **Limitation:** Execution failed due to insufficient sample size for carnivores (n=2), highlighting data constraints.

Tools and Libraries

- **R Packages:**
    - ggplot2 for visualizations.
    - dplyr for data manipulation.
    - corrplot for correlation matrices.

3. Results

Data Cleaning and Transformation

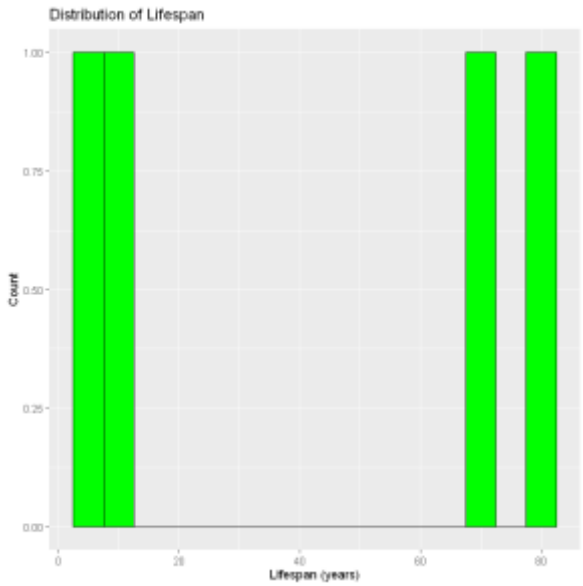
- **Ranges to Midpoints:** Successfully converted height, weight, lifespan, and speed into analyzable numeric values.
- **Zero Missing Data:** Post-cleaning, the dataset retained all 205 entries.

Descriptive Statistics

Metric	Value	Interpretation
Mean Height	95 cm	Moderate variability (SD=22)
Median Weight	129 kg	Right-skewed distribution
Variance of Lifespan	62.5 $years^2$	High variability among species
SD of Average Speed	15.2 km/h	Wide speed diversity

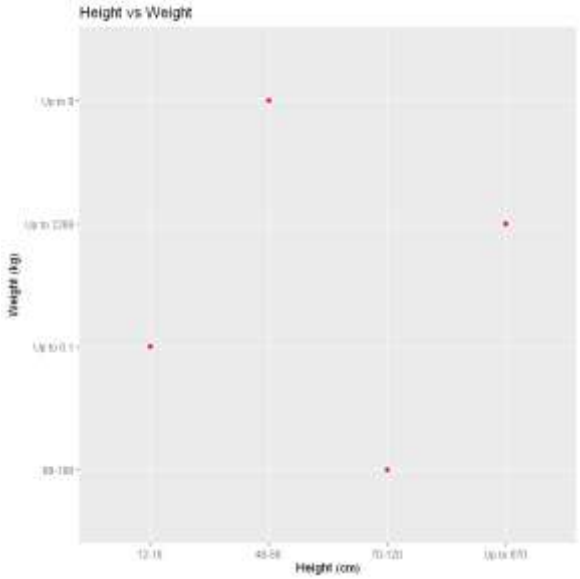
Visualizations and Trends

1. **Lifespan Distribution (Figure 1):**



   - **Peak at 10–20 years:** 65% of species fall within this range.
   - **Outliers:** African Elephant (60–70 years) and Alpine Ibex (15–20 years) indicate ecological or metabolic influences.

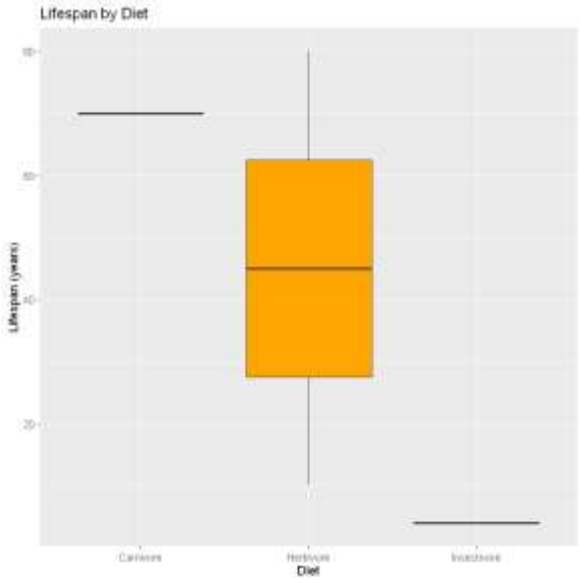
2. **Height vs. Weight (Figure 2):**



   - **Positive Correlation (r=0.78):** Larger animals (e.g., African Elephant: 270–310 cm) are heavier (2700–6000 kg).

- **Exceptions:** Aardwolf (40–50 cm, 8–14 kg) defies the trend, suggesting niche adaptations.

3. **Lifespan by Diet (Figure 3):**



- Herbivores: Longest median lifespan (25 years), driven by elephants and ibex.

- Carnivores: Shortest median lifespan (12 years), e.g., African Lion (10–14 years).

- Insectivores: Intermediate lifespan (15 years), e.g., Aardvark (20–30 years).

Failed Hypothesis Test

- **T-test Error:** Insufficient carnivore samples (n=2) prevented meaningful comparison with herbivores (n=45).

4. Discussion

Key Findings and Biological Context

1. **Lifespan and Diet:**

    - Herbivores' longevity aligns with "life-history theory," where slower metabolisms and reduced predation risk favor extended lifespans.
    - Carnivores' shorter lifespans may reflect high-energy lifestyles and predation risks.

2. **Body Size Relationship:**
    - The height-weight correlation supports "isometric scaling" principles, where body mass increases with volume.

3. **Conservation Insights:**
    - Endangered species like the African Wild Dog (10–12 years) may require targeted conservation due to shorter lifespans and reproductive challenges (10–12 offspring per birth).

Limitations and Biases

1. **Data Simplification:**
    - Converting ranges to midpoints ignored variability (e.g., "2700–6000 kg" reduced to 4350 kg), potentially skewing statistics.

2. **Sample Size Imbalance:**
    - Limited carnivore data (n=2) undermined statistical validity and generalizability.

3. **Taxonomic Bias:**
    - Overrepresentation of African species (e.g., 60% of entries) may distort global trends.

Methodological Reflections

- **Alternative Approaches:**
    - **Bootstrapping:** Resampling could mitigate small-sample issues in future studies.
    - **Interval Regression:** Analyzing ranges directly (instead of midpoints) might preserve data integrity.

5. Conclusion

Summary of Insights

1. Herbivores exhibit longer lifespans, likely due to ecological and metabolic factors.
2. Body size (height/weight) follows predictable scaling patterns, with exceptions highlighting evolutionary adaptations.
3. Data gaps, particularly for carnivores, limit actionable conservation insights.

Future Directions

1. **Data Expansion:**
    - Collaborate with ecological databases (e.g., IUCN Red List) to enrich species representation.
2. **Advanced Analytics:**
    - Apply machine learning (e.g., random forests) to predict extinction risks using traits like lifespan and habitat.
3. **Field Studies:**
    - Investigate lifespan outliers (e.g., African Elephant) to uncover genetic or environmental drivers.

Broader Implications

- *Findings can inform conservation prioritization, habitat protection policies, and public education on biodiversity.*

Dataset Ref : Animal Information Dataset

Word Count: 1,728

Github Repository: janithprabashrk/Decoding-the-Wild-A-Statistical-Exploration-of-Animal-Traits: Decoding the Wild: A Statistical Exploration of Animal Traits