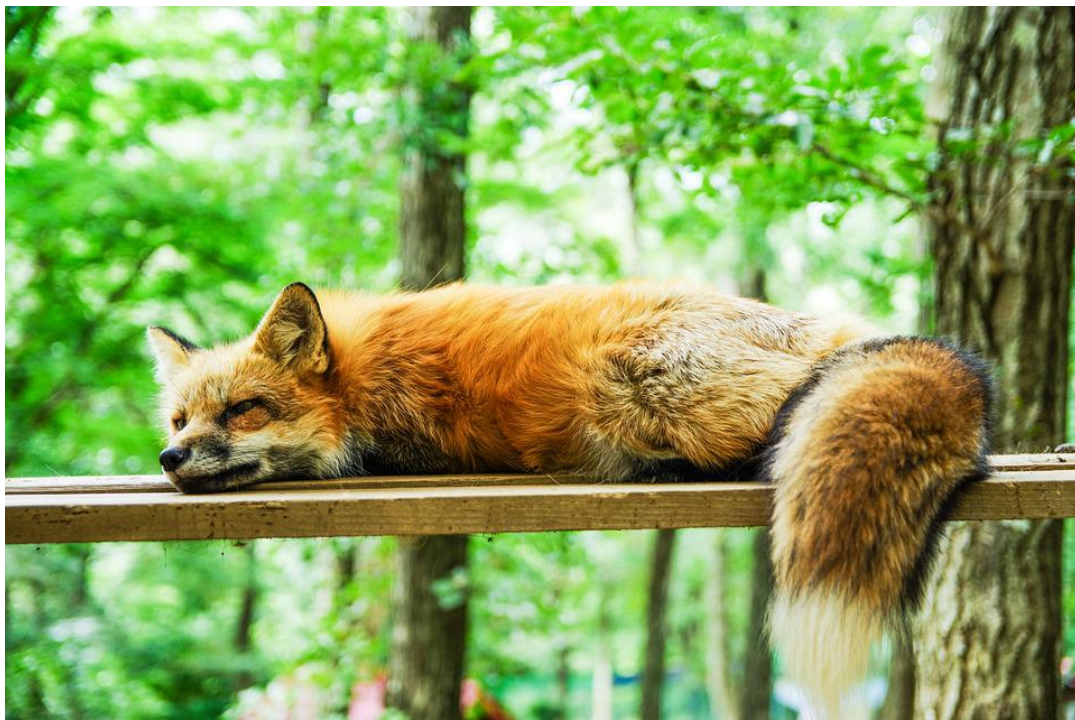**AUTONOMOUS UNIVERSITY OF BARCELONA**
COMPUTATIONAL MATHEMATICS AND DATA ANALYTICS

COMPLEX DATA ANALYTICS

A STATISTICAL APPROACH TO THE STUDY OF MAMMAL SLEEP PATTERNS IN RELATION TO BODY AND BRAIN WEIGHT

Pau Blasco Roca, NIU: 1600959

June 2022

Abstract

The intent of this paper is to study the relationship between the physical qualities of different species of mammals and their respective sleep patterns. These studies consist on an initial linear regression model, followed by a bootstrap approach. After that, further correlation and bootstrap studies are performed by separating mammals in several categories.

Introduction

Planet Earth has about ten million species of animals, each one of them with its own beautiful and intricate characteristics. This paper, however, will study one that cannot be seen or perceived directly: their sleep. Not only that, but this analysis will also study the length of their sleep cycles and the amount of time they stay in the REM phase.[1] This study will be performed over the "msleep"[2] dataset, which contains information about 83 different species of mammals, ranging from the Cheetah to the Mole Rat or the Red fox (the one in the paper's front page).

Initial observations

Before diving deep into the research, a previous study was performed, starting with a general analysis of the dataset and their respective graphic plots.
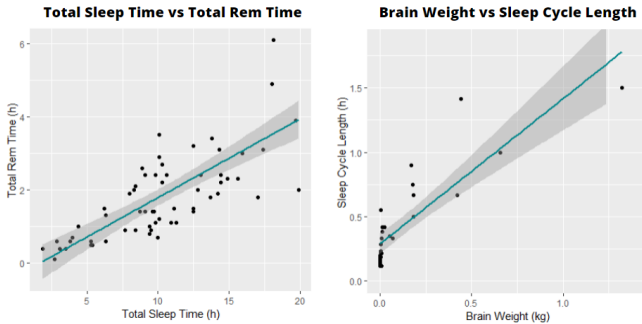

Figure 1: Initial comparisons between the variables of the dataset.

Total Sleep Time and REM Time Correlation

The Total Sleep Time (TST from now on) and the REM Time (RMT) appear to be positively correlated, with an adjusted R^2 of 0,56 and a proportionality coefficient of 2,62.

Brain Weight and Sleep Cycle Length Correlation

The Brain Weight (BRW from now on) and the Sleep Cycle Length (SCL) appear to be positively correlated, with an adjusted R^2 of 0,72 and a proportionality coefficient of 0,64.

Correlation Matrix   After observing these strong correlations, there was an interest in knowing whether there existed other dependencies between the variables in the dataset.

---

[1]Brown, Ritchie & McCarley, Robert. (2008). Neuroanatomical and neurochemical basis of wakefulness and REM sleep systems. Neurochemistry of Sleep and Wakefulness. 23-58.

[2]A quantitative, theoretical framework for understanding mammalian sleep. Proceedings of the National Academy of Sciences
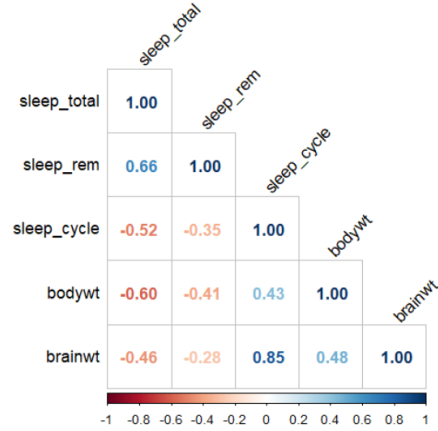

Figure 2: Correlation matrix of the dataset, computing the correlation between every pair of variables.

As it can be seen, there is also a strong negative correlation between BDW and TST, which might seem counter-intuitive since the natural supposition is that larger animals sleep more.

Parametric Bootstrap on initial paradigm

To obtain accurate results in a straightforward way, the Bootstrap Method[3] was applied. This method consists of resampling the dataset with replacement several times and achieves a more consistent and accurate result than a standard method would, as well as not needing the MLE to work.

These are the results after bootstrapping the dataset 10000 times with a sample of 35 values.

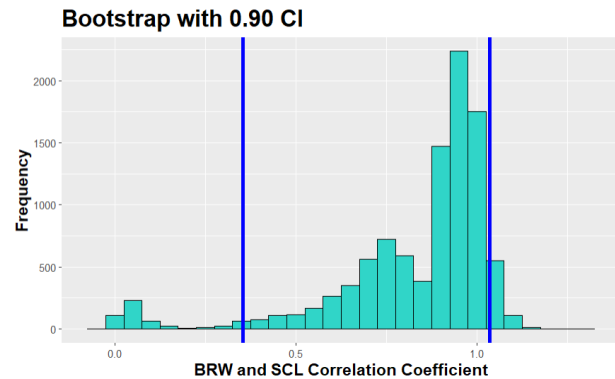### Brain Weight and Sleep Cycle Length P-Bootstrap

Figure 3: Histogram of the Bootstrap's analysis results with a 90% confidence interval

Here we obtain a mean correlation coefficient of 0.83 and a SE of 0.22. Our confidence interval, computed by the Bootstrap method, is $(0.35, 1.03)$. Furthermore, $0 \notin CI$, so we can affirm that BRW and SCL are *positively correlated*.

---

[3]Efron, Bradley.(1979). Bootstrap methods: another look at the jackknife.

Also, note that this interval contains the coefficient calculated by the Linear Regression Module, but overestimates it. This might be due to the Bootstrap Method not giving as much importance to outliers as the LRM.

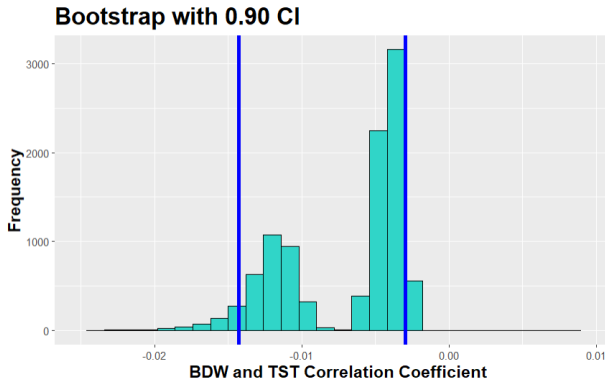### Body Weight and Total Sleep Time P-Bootstrap

Figure 4: Histogram of the Bootstrap's analysis results with a 90% confidence interval

In this case, we obtain a mean correlation coefficient of $-7.26 \cdot 10^{-3}$ and a SE of $5.74 \cdot 10^{-3}$. Our confidence interval, computed by the Bootstrap method, is $(-0.0143, -0.0030)$. Although the results obtained are close to zero, $0 \notin$ CI. This implies that BDW and TST are *negatively correlated*, at least, according to a Parametric Bootstrap Analysis.

Non-Parametric Bootstrap Analysis

After the latest analysis, a Non-Parametric Bootstrap Analysis was conducted. The data was supposed to be normally distributed (or to slightly follow a Normal Distribution). Using R's *boot* library, the two relationships (BRW vs SCL and BDW vs TST) were studied again, with the intention to have more meaningful and accurate results.

Brain Weight and Sleep Cycle Length NP-Bootstrap

After performing the NP-Bootstrap 10000 times, it was found that the correlation coefficient between the mammals' brain weight and their sleep cycle length was included in (0.7631, 0.9168), with a 90% confidence. This time, the results themselves were included in the interval previously computed by the P-Bootstrap analysis, narrowing it down.

Body Weight and Total Sleep Time NP-Bootstrap

After resampling the dataset 10000 times, the following confidence interval was found, with a 90% confidence interval: (-0.3798, -0.1458). These results indicate a much stronger correlation between the mammals' body weight and their total sleep time, given that the P-Bootstrap interval was much closer to zero. It must be taken into consideration that the two intervals (parametric and non-parametric) do not intersect.

Dataset divisions and new observations

Some of the results obtained were not precise enough, given the heterogeneity of our dataset. A way to narrow down the investigation was to subdivide the dataset into its three main groups: Carnivores, Omnivores and Herbivores.
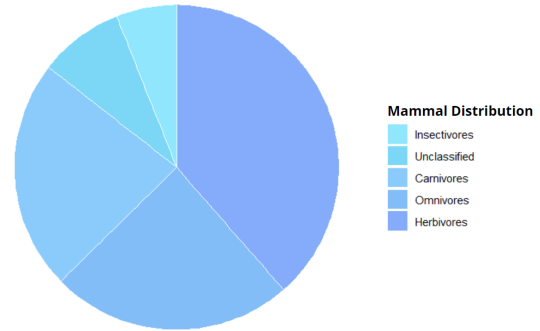

Figure 5: Distribution of the dataset based on mammals' diet

Insectivores and unclassified mammals were discarded, as they did not have enough data to perform Linear Regression and Bootstrap analysis. Even after this, as the dataset lacked significant data for the study, only detailed observations for specific groups could be made.

Differences between carnivores and herbivores

Notable differences between carnivores and herbivores were observed, as portrayed below. Meanwhile, omnivores did not seem to have any strong correlation, so they were excluded of the study.
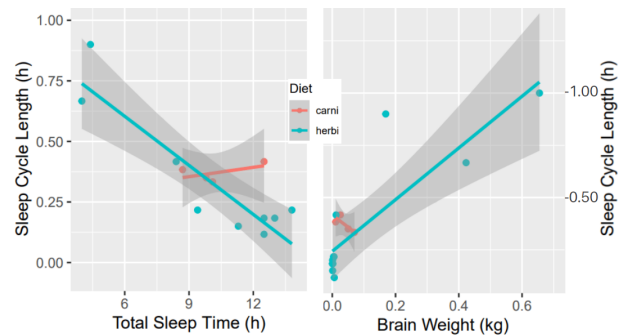

Figure 6: Linear regressions for both carnivore and herbivore mammals

Carnivores and herbivores appeared to have distinct correlation coefficients. Despite this, it cannot be sustained with certainty that their tendencies are different.

As observed in the figure on the left side, by removing the rightmost red point (which could likely be an outlier), all the carnivores fall into the herbivores' regression line. Meanwhile, in the figure on the right side, it can be seen that all the carnivores are already contained in the standard error area of the herbivore's linear regressor.

After this analysis and the non-disclosing results, we can hypothesize that, with more data entries, a clearer tendency might arise.

Bootstrapping herbivores and carnivores   Unfortunately, Bootstrap could not be applied here. The dataset's size was drastically reduced, and some entries were incomplete, thus further reducing its size. In some cases (i.e. carnivores' SCL) only up to 5 entries were available, making Bootstrap completely pointless in this case.

Conclusions

After performing several analyses on the dataset, and bootstrapping various combinations of variables, some points have been established, both partly confirming and expanding the bounds of our initial hypothesis.

- Mammals with a heavier brain have longer sleep cycles, with an estimate of 0.840 ± 0.076 hours of sleep cycle per kilogram of brain weight. This is the interval computed by the Non-Parametric Bootstrap, and it was included in the P-Bootstrap interval.

- Mammals with a heavier body have a shorter sleep time, with an estimate of -0.26 ± 0.12 hours of total sleep per kilogram of body weight. Note that the interval given by the NP-Bootstrap was accepted, but the interval computed by the P-Bootstrap method also showed a negative correlation with a 90% confidence interval.

- Differences based on their diet were observed, however the results were not sufficiently statistically significant to affirm said influence.

References

Brown, R., Ritchie & McCarley. (2008). Neuroanatomical and neurochemical basis of wakefulness and rem sleep systems. *Neurochemistry of Sleep and Wakefulness.*, 23–58.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist*, 1-26.

Van M. Savage, G. B. W. (2007). A quantitative, theoretical framework for understanding mammalian sleep. In *Proceedings of the national academy of sciences.* Los Alamos National Laboratory, NM: National Academy of Science of the United States of America.