The Duration of the Mammalian Sleep Cycle: A Linear Regression and Subsequent Analysis

Daria Stachowiak
Econometrics 272
Professor Woody Studenmund

Sleep serves as a necessary biological function in all living species on the planet, but sleep requirements vary drastically from species to species. A variety of benefits have been suggested for sleep, including benefits that are related to memory consolidation, energy conservation and the proper functioning of the immune system (Acerbi and Nunn 2005). By comparing sleeping habits to a number of species-specific characteristics that have resulted as an adaptation to different ecological niches, we can potentially explain why sleep is a crucial rather than an optional function for living organisms.

Through a study done by Zepelin and Rechtschaffen (Allison and Cicchetti 1976, 732), variables such as lifespan were correlated with certain sleep characteristics, supporting the theory that the data that will be presented in this analysis of 42 mammalian species is sufficient enough to BEGIN to try to understand the underlying connections between the duration of sleep of a species and various biological characteristics. On the other hand, Zepelin and Rechtschaffen isolated their study from ecological influences, summarizing their laboratory-based results by dividing mammals into "good sleepers" (more than 8 hours of sleep per day) and "poor sleepers" (less than 8 hours of sleep per day). The poor sleepers needed more time adapting to the laboratory setting to get a sound sleep, suggesting underlying environmental factors were in play. Species that are considered predators sleep longer than species that are typically preyed upon. In addition, species that normally sleep in enclosed spaces such as bats, sleep longer than sheep, who sleep in the open.

In this regression analysis, I will diverge into the next step and analyze the relationship between sleep duration and both constitutional characteristics AND ecological influences, finding them significant in predicting the amount of sleep that mammals get. I will be focusing solely on the duration of sleep rather than the time specifically spent in non-rapid eye movement (non-REM) sleep or rapid-eye movement sleep (REM), because this would require the study of multiple dependent variables and therefore, multiple regression analyses and the consideration of other potential independent variables that will not be discussed here.

After examining the data[1], I have decided to exclude certain mammalian species that have particularly unusual sleeping patterns such as dolphins, who shut down only half of their brains and opposite eyes when sleeping. I have also decided to not take into consideration data for the very few mammalian species that primarily use their hind limbs , because they are likely to exert more energy when they move and this could potentially affect their sleep duration (Gonfalone 2016). Such precautions and exclusions were taken in order to minimize as much impact on the data as possible by outside variables that could not be accounted for.

For my regression analysis of mammalian duration of sleep, I will use numerical data collected and provided by Zepelin and Rechtschaffen with the addition of my own independent variable of Diet, for which I had created a variable based on what the majority of a species' diet consists of. I assigned an appropriate number of 1-3 to each of the 42 mammalian species based on my criteria.

[1] Appendix 1

THE VARIABLES

DEPENDANT VARIABLE :

VARIABLE	Description
TotalSleep	The duration of sleep of a mammalian species measured in hours per day

INDEPENDENT VARIABLES:

VARIABLE	Description
**BodyWt	Body weight measured in kilograms (kg)
BrainWt	Brain weight measured in kilograms (kg)
LifeSpan	Maximum life span measured in years (yrs)
Gestation	Gestation time measured in days (days)
Diet	Diet Index: 1=carnivore 2=omnivore 3=herbivore

Danger	Overall Danger Index: Sleep Exposure Index (1-5) 1=least exposed (e.g. animal who is sound asleep in a well-protected den) 5=most exposed (e.g. animal who sleep in open field with no protection)(1-5) Predation Index (1-5) 1=least likely to be preyed upon; 5=most likely to be preyed upon This variable utilizes both of these factors in determining overall danger. For example, a species may have a Sleep Exposure Index of 5, but due to not encountering any risk from other animals or environmental factors, may have have a Danger Index of 1

**Not utilized in regression

****BODY WEIGHT:**

It is important to take into consideration brain weight and body weight without much distinction between them because the brain and body weight together could account for 31% of variance in a mammal's sleep cycle (Zepelin and Siegel and Tobler 2005). If one takes into account the Energy Conservation Theory, species that are greater in mass endure less pressure to sleep because they have greater reserves of energy. As an animal increases in mass, the proportion of fat to body mass also increases. Consequently, it is not surprising that smaller animals may have fat reserves as low as 5% while species that weigh up to 1000kg could potentially have a fat to body mass ratio of 25-30% (Allison and Chicchetti 1976).

To further elaborate, an animal that is larger in size has a lower thermal conductance, meaning they lose heat from their bodies much more slowly than smaller animals. As a result, they do not have to actively produce as much heat and it is easier for their body to maintain homeostasis. At thermoneutrality, the survival time of a much smaller animal may be a low as a couple of hours and this may be especially emphasized in colder temperatures. Supporting this conservation theory is the long sleep cycle of mammals that are only beginning to enter maturation (Zepelin and Siegel and Tobler 2005). During this period of time, growth requires immense amounts of energy. I hypothesize, using this data, that there will be **a statistically significant negative relationship between body weight and duration of sleep in a species.**

BRAIN WEIGHT:

Brain weight and sleep cycle length have been found to be correlated independently, but body weight did not correlate independently of brain weight. Therefore, it is important to consider the combined weight of the body and the brain (Zepelin and Siegel and Tobler 2005). Since there is an extreme correlation between Brain Weight and Body Weight in this set of data ($r = .9885$), it is best to utilize Brain Weight as independent variable in the equation in place of Body Weight, to minimize potential error and simultaneously showcase the relationship between Body Weight and duration of sleep. Brain also seems to be a better predictor than body weight because there is less variation in brain weight. I hypothesize that there will be **a statistically significant negative linear relationship between brain weight and duration of sleep in a species.**

LIFESPAN:

For 27 species of mammals, a study was done concerning the variation sleep duration between the species. To formulate a potential connection between sleep duration and the life span of a species, the median life span of the species was calculated =17.5 years. The species were then broken into two groups- those living less 17.5 years and those living more than 17.5 years. When comparing, the mean hours of sleep per day for animals with the shorter life span was 12.5, while the mean hours of sleep per day for animals with a longer life span was 5.5, suggesting a 7 hour on average difference between longer living and shorter living species ("Distribution of Sleeping Times"). I hypothesize that there will be **a statistically significant negative linear relationship between lifespan and sleep duration of a species.**

GESTATION:

The shorter the period of gestation, the less time for maturation of the fetus. As a result, the baby needs to be protected for longer and therefore their sleep time is increased (Siegel 2005) . Altriciality, or the incapability of functioning on one's own when born may cause animals to sleep for a longer period of time in order to not only fully develop their sensory organs, but also to strengthen their adaptation to the external senses that they experience such as light, sound, taste, odor as well as fluctuations in temperature and the the gravitational pull (Gonfalone 2016) . REM sleep is the primary component of a altricial animal's sleep cycle because during this stage, the brain can process the experiences it had when awake and utilize this information to enhance the next period of awakeness. I hypothesize that there will be **a statistically significant negative relationship between gestation period and time spent sleeping.**

DANGER:

Zepelin and Rechtschaffen utilized an index ranging from 1-5 in order to rate potential dangers posed to a species, taking into account both Predation and Exposure. Since this variable took both of these independent variables into account in order to create an overall evaluation, it is a better assessment of the relationship between a species' necessary actions for survival and its duration of sleep (Allison and Cicchetti 1976).

A study was done that examined different scenarios of predation pressure. In one of the scenarios, predators exhibited predictable and single-phase behaviour and "potential prey" at sleep sites were expected to perfectly safe from the risk of predation. As expected, the prey behaved asynchronously to predators. In another scenario, predators had the same fixed single-phase activity behaviour only the safety of the sleep site of "potential prey" varied from perfect safety to extreme risk. Prey adjusted their circadian rhythm relative to how safe their sleep sites were as well as the time of day of their predator activity. Variable predation risk also had an impact on sleep duration in prey. If they were considered safer when sleeping, they slept more when predators were active and they foraged during the other part of the day. When they were safer when awake, they shifted their activity to include a greater proportion of foraging each day (Acerbi and Nunn 2005). Predators and prey are interdependent and manage to adjust their schedules to minimize their risk and maximize their food finds. I hypothesize that there will be **a particularly strong negative linear relationship between danger and duration of sleep.**

DIET:

In most cases, herbivores sleep less than carnivores because of the longer amount it time that is needed for foraging and consumption of enough calories to sustain themselves. Species that are near the top of the food chain such as tigers, have little fear of other animals and may eat fulfilling meals at one time, once they catch their prey. They are able to spend a greater majority of their day sleeping (Mastin). Due to these findings, I hypothesize that there will be **a significant negative linear relationship between the diet of a species and their duration of sleep.** Species that are primarily herbivores will forage more and sleep less, while species that are primarily carnivores, will have fewer predators and be able to sleep more.

Regression Analysis

Regression I[2]

For the first run of my regression, I decided to include the variables BRAINWT, MAXLIFESPAN, GESTATION, DANGER and DIET due to overwhelming support from scientific research and the subsequent theories. I decided to omit BODYWT permanently because the correlation between BRAINWT AND BODYWT was r = .9885, meaning including both would simply be redundant. I favored BRAINWT over BODYWT due to the smaller variation in BRAINWT.

The results were as follows:

$$\widehat{TOTAL\,SLEEP}_i = \underset{}{17.89707} + \underset{\substack{(1.067032)\\ t = 1.37}}{1.458927BRAINWT_i}$$

$$- \underset{(.0355692)}{.0415886MAXLIFESPAN_i} - \underset{(.0069231)}{.0189701GESTATION_i}$$

[2] Appendix 1

t = -1.17 -2.74

- 1.593805DANGER$_i$ - .0344918DIET$_i$

(.4632179) (.7994513)

t = -3.28 -0.15

N = 42 $\overline{R^2}$ = 0.5266

t-Test:

In order to determine whether all of the variables are statistically significant and/or if the coefficients are in the expected direction, I will conduct a one-sided t-test for all five independent variables at a 5% level of significance. Since N = 42, K = 5, N - K - 1 = 36

For 36 degrees of freedom, the critical t-value or t_c is 1.684 (rounded up to 40 degrees of freedom).

BRAINWT$_i$:

$H_0 = B \leq 0$

$H_A = B > 0$

t = 1.37

$t_c > |1.37|$

Although the coefficient of BRAINWT$_i$ is in the expected (+) direction, the null hypothesis cannot be rejected because the critical value of t is greater than the $|t - score|$ of BRAINWT$_i$, meaning this variable is not as statistically significant in this equation, contrary to my hypothesis.

MAXLIFESPAN$_i$:

$H_0 = B \geq 0$

$H_A = B < 0$

t = - 1.17

$t_c > |-1.17|$

Although the coefficient of MAXLIFESPAN_i is in the expected (-) direction, the null hypothesis cannot be rejected because the critical value of t is greater than the $|t - score|$ of MAXLIFESPAN_i, meaning this variable is not statistically significant in this equation, contrary to my hypothesis.

GESTATION_i :

$H_0 = B \geq 0$

$H_A = B < 0$

t = - 2.74

$t_c < |-2.74|$

The coefficient of GESTATION_i is in the expected (-) direction and the critical value of t is less than the $|t - score|$ of GESTATION_i, meaning this variable is statistically significant in this equation at a 5% level of significance and the null hypothesis can be rejected.

DANGER_i :

$H_0 = B \geq 0$

$H_A = B < 0$

t = - 3.28

$t_c < |-3.28|$

The coefficient of DANGER_i is in the expected (-) direction and the critical value of t is less than the $|t - score|$ of DANGER_i, meaning this variable is statistically significant in this equation at a 5% level of significance and the null hypothesis can be rejected.

DIET_i:

$H_0 = B \geq 0$

$H_A = B < 0$

t = - 0.15

$t_c > |-0.15|$

Although the coefficient of DIET_i is in the expected (-) direction, the null hypothesis cannot be rejected because the critical value of t is immensely greater than the $|t - score|$ of DIET_i, meaning this variable is not at all statistically significant in this equation, contrary to my hypothesis.

Omitted Variable (Violation of Classical Assumption III):

While there is always a possibility of a significant explanatory variable being omitted from this equation, it does not seem likely that there is bias due to the fact that all of five of the coefficients of the independent variables in the equation were found to be in the expected direction. In fact, the variables that have shown the strongest statistical significance such as GESTATION, have coefficients that are theoretically sound in terms of size. Holding all other independent variables in the equation constant, when the gestation period increases by one day, sleep duration decreases by .0415886 hours.

Irrelevant Variable

An irrelevant variable will not cause bias in the coefficients of the other independent variables, but it may potentially increase the variance of the of the estimated coefficients. In this regression, the variable, DIET, had a coefficient in the expected direction (-), however, the t-score was exceptionally low at -0.15. This made me decide to reexamine my initial theory of diet having an impact on the duration of sleep of animals. Although it is reasonable to conclude that mammals that spend more time scavenging potentially sleep less and that carnivores spend less time searching for food due to their immense intake of their needed energy for the day through relatively few hunts throughout the day, it is difficult to classify omnivores. Omnivores are scattered throughout the food-chain, therefore their time spent sleeping/hiding and time spent looking for food varies depending on the species. It is also quite possible that on some days, omnivores may spend a greater or lesser portion of their day on obtaining food depending on what is available, whether this is prey or vegetation. It is also quite possible that this variable is not necessary in the equation because the characteristics of herbivores, omnivores and carnivores are already accounted for by other variables in the equation. Metabolism, which determines the amount of food an animal needs, is associated with body weight and body weight is extremely correlated with brain weight. The danger index also takes into account an animal's eating habits because carnivores tend to have fewer presented dangers when compared to herbivores. For these reasons, the DIET variable that I have included, which placed dependence on whether a species is an herbivore, omnivore or carnivore, will be omitted from my next regression.

Although MAXLIFESPAN and BRAINWT also had t-scores that were statistically insignificant, they were relatively close to significance. Both theories for keeping the variables in the equation are too strong to overlook.

Incorrect Functional Form

This equation was specified linear functional form:

$$Y_i = B_o + B_1 X_i + \varepsilon_i$$

There is little reason to doubt the use of the linear functional form for this regression because research supports the hypotheses that there are linear relationships between each of the independent variables specified in the equation (holding all other independent variables constant!) and the dependent variable. As biological development occurs along a linear timeline, it is reasonable to deduce, for example that the longer gestation occurs, the more mature the animal is when they are born, the less time they need for sleep in order to further develop. For the danger index, for example, the level of danger an animal encounters is relatively proportional to the amount of sleep it gets. All of the variables included can be explained in a linear fashion.

Multicollinearity (If perfect, it is a violation of Classical Assumption VI)

If multicollinearity is a concern in an equation, the estimates of the coefficients will still remain unbiased, but it is possible that the variances and standard errors will increase, and if a variable is dropped from an equation, drastic changes in the values of the estimates can occur. T-scores tend to also be low if multicollinearity is present. If the independent variables that are found to be multicollinear continue to behave in the same predicted manner in the future, than predictions utilizing the equation will not be affected. The overall fit will also not be affected.

To test for collinearity in my equation, I found the correlation coefficients between $BRAINWT_i$, $MAXLIFESPAN_i$, $GESTATION_i$, $DANGER_I$ and $DIET_i$.

None of the variables showed correlation > .8 , with the highest being a correlation coefficient of .7337 between $GESTATION_I$ and $BRAINWT_i$ [3].

Variance Inflation Factor[4]:

The Variance Inflation Factor shows us how much multicollinearity has increased the variance of an estimated coefficient (Studenmund 234). Therefore, it is beneficial if the VIF is as low as possible. (APPENDIX) A VIF of 5 signifies a case of severe multicollinearity. Although the mean VIF for the equation is 2.16 and no VIF is higher than 3.06 for $GESTATION_I$, multicollinear effects could have still played a part in the outcome of the equation. Therefore, we cannot reject the possibility of multicollinearity.

---

[3] Appendix 2
[4] Appendix 2

Regression II[5]:

$$\hat{TOTAL\ SLEEP}_i = 17.75924 + 1.465801 BRAINWT_i$$
(1.051864)
t = 1.39

$- .0408068 MAXLIFESPAN_i - .0192515 GESTATION_i$
(.0347152) (.0065741)
t = -1.18 -2.93

$- 1.554141 DANGER_i$
(.3910188)
t = -3.97

N = 42 $\overline{R^2} = 0.5391$

t-Tests:

In order to determine whether all of the variables are statistically significant and/or the coefficients are in the expected direction, I will conduct a one-sided t-test for all five independent variables at a 5% level of significance. Since N = 42, K = 4, N - K - 1 = 37

[5] Appendix 3

For 37 degrees of freedom, the critical t-value or t_c is 1.684 (rounded up to 40 degrees of freedom).

BRAINWT$_i$:

$H_0 = B \leq 0$

$H_A = B > 0$

t = 1.39

$t_c > |1.39|$

Although the coefficient of BRAINWT$_i$ is in the expected (+) direction, the null hypothesis cannot be rejected because the critical value of t is greater than the $|t - score|$ of BRAINWT$_i$, meaning this variable is not as statistically significant in this equation, contrary to my hypothesis.

MAXLIFESPAN$_i$:

$H_0 = B \geq 0$

$H_A = B < 0$

t = - 1.18

$t_c > |-1.18|$

Although the coefficient of MAXLIFESPAN$_i$ is in the expected (-) direction, the null hypothesis cannot be rejected because the critical value of t is greater than the $|t - score|$ of MAXLIFESPAN$_i$, meaning this variable is not statistically significant in this equation, contrary to my hypothesis.

GESTATION$_i$:

$H_0 = \ B \geq 0$

$H_A = B < 0$

t = - 2.93

$t_c \ < \ |-2.93|$

The coefficient of $\text{GESTATION}_\text{i}$ is in the expected (-) direction and the critical value of t is less than the $|t - score|$ of $\text{GESTATION}_\text{i}$, meaning this variable is statistically significant in this equation at a 5% level of significance and the null hypothesis can be rejected.

DANGER_i:

$H_0 = \ B \geq 0$

$H_A = B < 0$

t = - 3.97

$t_c \ < \ |-3.97|$

The coefficient of DANGER_i is in the expected (-) direction and the critical value of t is less than the $|t - score|$ of DANGER_i, meaning this variable is statistically significant in this equation at a 5% level of significance and the null hypothesis can be rejected.

Comparison of Regression Equation I and II:

1.) Theory: While the inclusion of the DIET variable in equation I, made sense on a surface level, the faults with this variable could simply not be overlooked. The theory behind it, as explained above (CITE), renders it insignificant because it can be accounted for in different variables and it is also measured in a way that would likely lead to ambiguous results. Therefore, theory supports equation II.

2.) t-Test: While the variables of MAXLIFESPAN and BRAINWEIGHT are still statistically insignificant when tested at the 5% level and at 40 degrees of freedom, their t-scores did increase when DIET was dropped and they are significant at SOME levels. The t-Tests support equation II.

3.) $\overline{R^2}$: $\overline{R^2}$ increased from .5266 to .5391, supporting equation II.

4.) Bias: The coefficients for the independent variables each changed by less than a single standard error, suggesting that there was barely any bias caused by including the variable and further supporting the theory of its irrelevance.

Taking the above factors into consideration, there is evidence that equation II is the better fit of the two attempted regression equations.

Proceeding with Equation II:

Heteroskedasticity (Violation of Classical Assumption V):

The presence of heteroskedasticity signifies that the variability of a variable is unequal across the range of values of a second variable that predicts it. When heteroskedasticity is present, Ordinary Least-Squares is no longer the minimum variance estimator. Heteroskedasticity can be pure and caused by the error term of an equation or impure, which may be a sign that the equation is incorrectly specified in some way.

Testing for heteroskedasticity[6]:

[6] Appendix 3

To test for heteroskedasticity, we can use the Breusch-Pagan test, which lets us determine if we can explain the squared residuals with the proportionality factors.

After running a regression in STATA with the squared residual as the function of the independent variables in the equation…

$$e^2{}_i = \alpha_0 + \alpha_1 BRAINWT_{1i} + \alpha_2 MAXLIFESPAN_{2i} + \alpha_3 GESTATION_{3i} + \alpha_4 DANGER_{4i} + u_i$$

We can hypothesize:

$H_0 = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

$H_A = H_0 \; is \, false$

$R^2 = .2232$

$N = 42$

$Degrees \; of \; Freedom = 4$

$NR^2 = 9.3744$

Critical chi square value at 5% *level of significance* = 9.49

Since NR^2 is not greater than the chi-square value, we cannot reject the null hypothesis of homoskedasticity. Therefore, heteroskedasticity in the equation is not confirmed.

Serial Correlation (Violation of Classical Assumption IV):

Pure serial correlation can occur in time-series data if the stochastic error term in an equation is influenced by events that took during a previous period. However, cross-sectional data may test positively for impure serial correlation if there is a likelihood that the equation wasn't correctly specified or a relevant variable was omitted or incorrectly specified.

The Durbin-Watson Test[7]

To test for first-order serial correlation, I had STATA recognize my cross-sectional equation as a time-series equation.

Hypotheses:

$H_0: \ P \leq 0 \ (No\ positive\ serial\ correlation)$

$H_A: \ P > 0 \ (Positive\ serial\ correlation)$

$K = 4$

$N = 42 \ (rounded\ to\ 40)$

$d_L = 1.29$ at 5% one-sided level of significance

$d_U = 1.72$ at 5% one-side level of significance

Generated d-value = 1.141605

Since $d < d_{l,}$ we reject the null hypothesis of no positive serial correlation.

This outcome is a bit troublesome considering the fact that we are dealing with cross-sectional rather than time-series data. Therefore, this is impure serial correlation and may be a sign of some greater issues with the specification of our equation. Although typical serial correlation can be virtually undetectable since it does not cause any bias in the estimates of our coefficients, it DOES lead to inaccurate t-scores due because it typically underestimates the Standard Errors. Additionally, Ordinary-Least Squares is no longer a minimum variance estimator.

Conclusion:

[7] Appendix 3

To conclude my regression analysis of the sleep duration of mammals, I was able to explain the majority of the variance in sleep times within my data sample by taking into consideration the weight of a mammal's brain, their max lifespan (on average), their period of gestation, and the level of danger their incur on a daily basis whether this is due to the exposure they incur from their sleeping habits or their likelihood to get preyed upon. As evidenced, Danger and Gestation period have the most significant impact on a mammalian species' amount of sleep. Nevertheless, the subject of sleep is a very complex one that requires a much more extensive analysis than the overview in this paper. Factors such as the differences in types of sleep that species undergo as well as as varies physical characteristics that differentiate certain mammalian species need to be considered before the reason and purpose for such a complex biological function can be established.

After finding impure serial correlation in my final regression equation and not being able to simply remedy this by utilizing Newey-West Standard Errors, I have begun to realize that there may have been important explanatory variables that could have provided an even better fit for this equation that I simply hadn't even considered in the first place. It is entirely possible that my data set is too small to try to summarize one of the most complex classes of animals on the planet and this could have potentially led to some unknown biases. There could have been other environmental factors besides danger that should have possibly been taken into account such as whether a species is aquatic or terrestrial. The subject leaves much to be desired, but a solid regression is a good start at explaining duration of sleep in mammalian species.

Bibliography

Acerbi, Alberto and Charles L. Nunn. “Predation and the phasing of sleep: An evolutionary individual-based model.” *Animal Behaviour* (2005) 81(4): 801-811.

Allison, T., and D. Cicchetti. "Sleep in Mammals: Ecological and Constitutional Correlates." *Science* 194.4266 (1976): 732-34. Web.

Al-Majed, N.b, and M.r Preston. "Factors Influencing the Total Mercury and Methyl Mercury in the Hair of the Fishermen of Kuwait." *Environmental Pollution* 109.2 (2000): 239-50. Web.

Gonfalone, Alain. "Negative Correlation between Gestation and Sleep Durations in Mammals." *Open Access Animal Physiology* Volume 8 (2016): 1-7. Web.

Mastin, Luke. "Sleep Anthropology - Sleep in the Animal Kingdom." *How Sleep Works*. N.p., n.d. Web.

Siegel, Jerome M. "Clues to the Functions of Mammalian Sleep." *Nature* 437.7063 (2005): 1264-271. Web.

Studenmund, A. H. *Using Econometrics: A Practical Guide*. Boston, MA: Pearson, 2016. Print.

Zareba, G., E. Cernichiari, and LA Goldsmith. "Validity of Methyl Mercury Hair Analysis: Mercury Monitoring in Human Scalp/nude Mouse Model." *Journal of Applied Toxicology : JAT.* U.S. National Library of Medicine, 2007. Web.

Zepelin, Harold, Jerome Siegel, and Irene Tobler. "Mammalian Sleep." *Mammalian Sleep* (2005): 91-99. 1 June 2005. Web.

"Distribution of Sleeping times." *BGSU*. N.p., n.d. Web.