Introduction to Statistics

Final problem set // December 6, 2024

Dylan Dettloff, Alison Lawyer, Laura Mitterrand

Analysis 1: Relationship Between Brain Size and Dreaming Sleep Ratio

Rationale and Hypothesis:

Hypothesis: Species with larger brain-to-body weight ratios will spend a higher proportion of their sleep dreaming. This is based on the idea that dreaming (paradoxical sleep) may support brain functions like memory consolidation and complex cognition, which are associated with larger brains.

Data Preparation:

- Cleaned the dataset (mammals_sleep == -999 converted to NA).
- Removed rows with missing values.
- Created a new column dream_sleep_ratio as the percentage of dreaming sleep relative to total sleep.

Exploratory Analysis:

Descriptive statistics for the variables:

Variables	Overall (N=48)
brain_weight_body_weight_percent	
Mean (SD)	1.11 (0.944)
Median [Min, Max]	0.740 [0.0910, 3.96]
dream_sleep_ratio	
Mean (SD)	18.6 (8.95)
Median [Min, Max]	17.6 [0, 46.2]

The descriptive statistics for brain size reveal considerable variability, with a mean of 1.11189 and a standard deviation of 0.9441679. The wide range from 0.09097 to 3.96040 suggests significant diversity in brain size across mammalian species. For dreaming sleep percentage, the mean of 18.65% indicates that mammals typically spend about one-fifth of their sleep time in paradoxical sleep, with substantial variation (standard deviation of 8.95).

Main Analysis:

Linear regression: lm(dream_sleep_ratio ~ brain_weight_body_weight_percent).

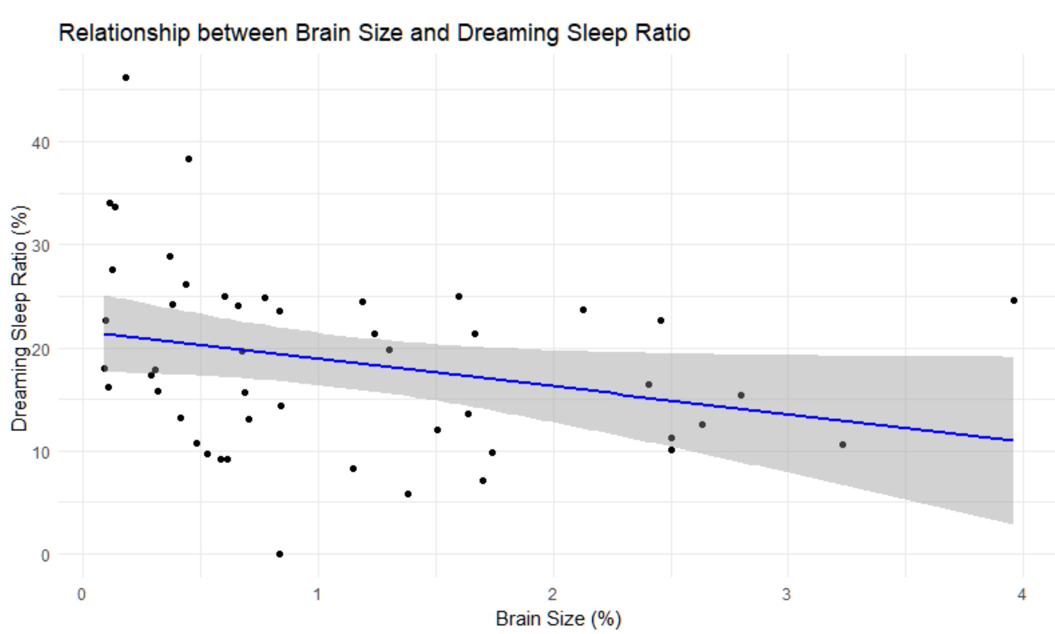
- Residual standard error: 8.679 on 46 degrees of freedom
- Multiple R-squared: 0.08011,Adjusted R-squared: 0.06012
- F-statistic: 4.006 on 1 and 46 DF, p-value: 0.05125
- The linear regression results show a weak relationship between brain size and dreaming sleep, with a low R-squared of 0.08011, suggesting that brain size explains only about 8% of the variation in dreaming sleep. The p-value of 0.05125 is significant, indicating a potential connection.

Non-linear regression: Added brain_size_sq as a predictor in a quadratic model.

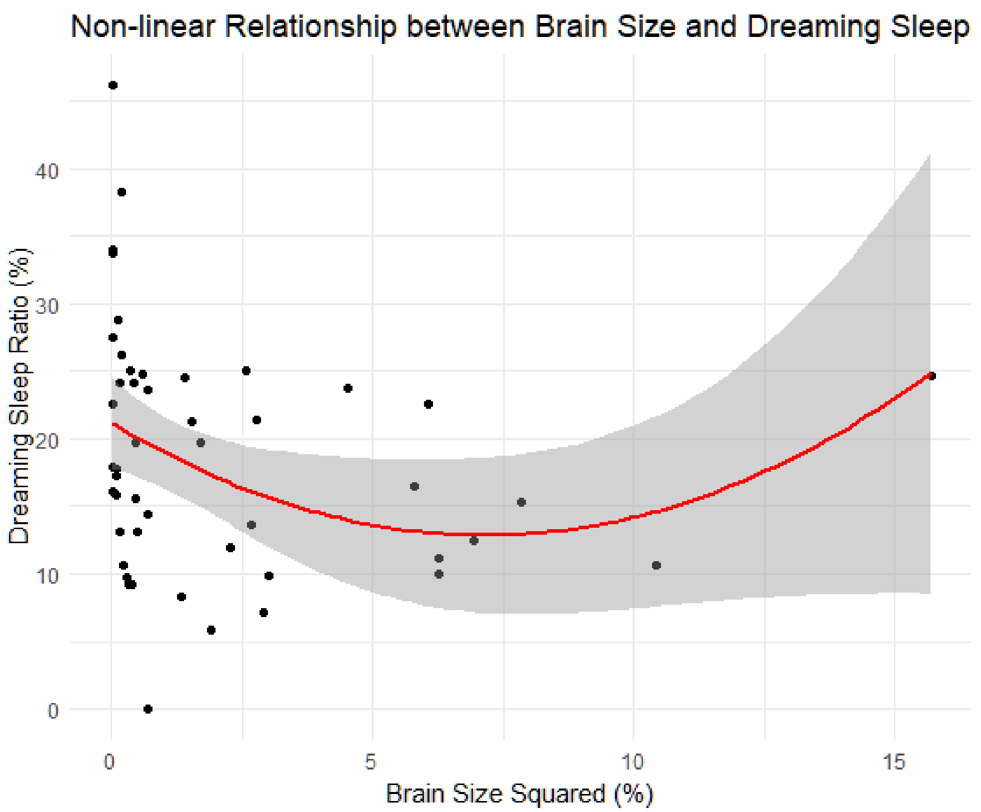
- Residual standard error: 8.247 on 45 degrees of freedom
- Multiple R-squared: 0.1875, Adjusted R-squared: 0.1514
- F-statistic: 5.191 on 2 and 45 DF, p-value: 0.009364
- The non-linear quadratic model slightly improved the explanatory power, increasing the R-squared to 0.1875 and producing a significant result (p-value = 0.009364). This suggests that the relationship between brain size and dreaming sleep might be more complex than a simple linear association.

Visualization:

Scatterplot with a linear regression line:



Quadratic plot for non-linear trends:



Discussion:

If the hypothesis is supported, larger brain-to-body weight percentages will correlate with higher dreaming-sleep ratios. However, the hypothesis is not supported. This analysis has shown no significant relationship between Brain Size (proportional to body size) and Paradoxical Sleeping time (proportional to total sleeping time).

Analysis 2: Predator Exposure and Dreaming Sleep

Rationale and Hypothesis:

Hypothesis: Species with low predator exposure spend more time dreaming due to increased safety. Dreaming may serve functions like memory consolidation, which could be deprioritized under high-risk conditions.

Data Preparation:

- Recoded sleep_exposure_index into a binary variable predator_status:
  - ≤2 = "Low Exposure," >2 = "High Exposure."
- Calculated dream_sleep_ratio.

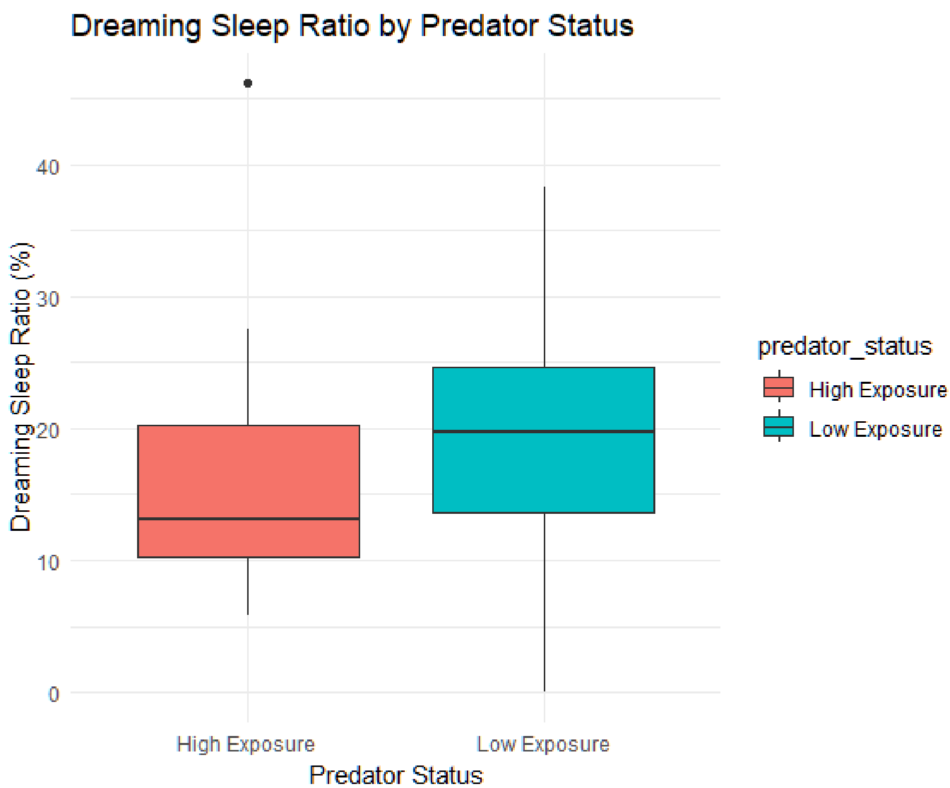
Exploratory Analysis:

Summary statistics for dream_sleep_ratio across predator status groups.

Variables	High Exposure (N=15)	Low Exposure (N=33)	Overall (N=48)
brain_weight_body_weight_percent			
Mean (SD)	0.740 (0.762)	1.28 (0.980)	1.11 (0.944)
Median [Min, Max]	0.416 [0.0910, 2.63]	0.833 [0.111, 3.96]	0.740 [0.0910, 3.96]
dream_sleep_ratio			
Mean (SD)	16.7 (10.3)	19.6 (8.26)	18.6 (8.95)
Median [Min, Max]	13.2 [5.83, 46.2]	19.8 [0, 38.3]	17.6 [0, 46.2]

The summary statistics for dream sleep ratio across predator status groups show interesting nuances. High-exposure species have a lower mean dream sleep ratio (16.7%) with greater variability (SD = 10.3), while low-exposure species demonstrate a slightly higher mean (19.6%) with less variation (SD = 8.26).

Boxplots for visual comparisons:



Main Analysis:

T-test to compare dream_sleep_ratio between predator exposure groups.

- t = -0.95773, df = 22.467, p-value = 0.3484
- The t-test results (t = -0.95773, df = 22.467, p-value = 0.3484) indicate no statistically significant difference in dream sleep ratios between high and low predator exposure groups. This suggests that predator exposure may not directly influence the proportion of time spent in paradoxical sleep, contrary to the initial hypothesis.
- The lack of statistical significance does not necessarily mean no relationship exists but rather that any potential relationship is not strong enough to be detected with the current sample and methodology.

Discussion:

If supported, species in "Low Exposure" will show significantly higher dreaming sleep ratios compared to "High Exposure." The hypothesis is not supported. This analysis shows no significant relationship between predator status and dreaming sleep ratio.

Analysis 3: Predation index effect on total sleep

Rationale and Hypothesis:

The dataset includes data describing mammals' total sleep time in hours per day as well as their predation levels, 1 being mammals experiencing the least amount of predation.
Hypothesis: Mammals subject to the least amount of predation (i.e. group 1 and 2 of the predation index) have a higher total sleeping time per day. Sleep is necessary to all mammals, yet it is a highly vulnerable state and from an evolutionary perspective it makes sense that those mammals subject to many predators that sleep less would have been fittest and selected for, whereas mammals that don't have to constantly stay alert for signs of danger wouldn't have the pressure of selection for shorter sleeping periods.

Data Preparation:

- Renamed the columns we are concerned with to SWS and PS.
- Recoded the predation_index variable into a factor.

Exploratory Analysis:

Since we have 5 groups (predation index 1 - 5) as categorical variables (in effect ordinal variable) with 7 to 15 observations in each, and one continuous variable (total_sleep) we will proceed to the following exploratory analysis: summary of central tendency and variances of total sleep for each predation index.

Variables	1 (N=14)	2 (N=15)	3 (N=12)	4 (N=7)	5 (N=14)	Overall (N=62)
total_sleep						
Mean (SD)	12.1 (4.60)	12.7 (3.93)	9.12 (4.53)	10.2 (2.44)	7.38 (4.81)	10.5 (4.61)
Median [Min, Max]	11.4 [6.10, 19.9]	11.0 [6.60, 19.4]	9.30 [3.30, 15.8]	10.3 [6.20, 13.2]	6.05 [2.60, 16.5]	10.5 [2.60, 19.9]
Missing	0 (0%)	0 (0%)	2 (16.7%)	0 (0%)	2 (14.3%)	4 (6.5%)

Table 1: Central tendency, variance and confidence interval for total sleep by predation level.

Bar graph: We can see a general trend of a decrease in total sleep as predation increases, however there's a lot of overlap in the 95% confidence interval areas.

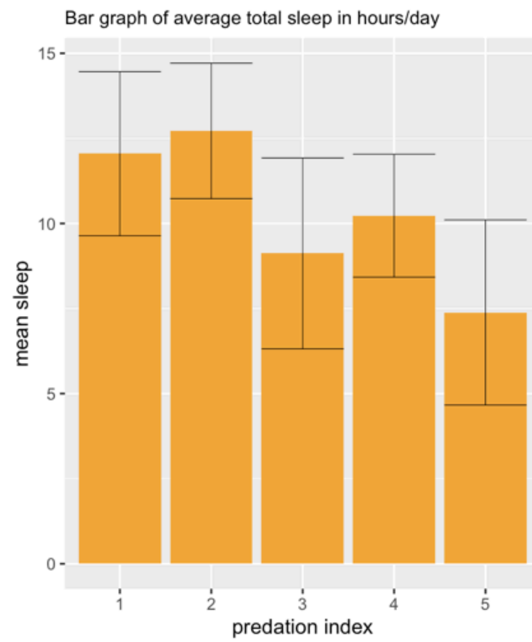


Figure 1: Graph of the average total sleep of mammals for each predation index, showing the 95% confidence interval.

Boxplot: We can visualize the measure of spread of the total sleep time for each predation index. We confirm the overlapping between index 1 and 2, and the general tendency seen in the bar graph.

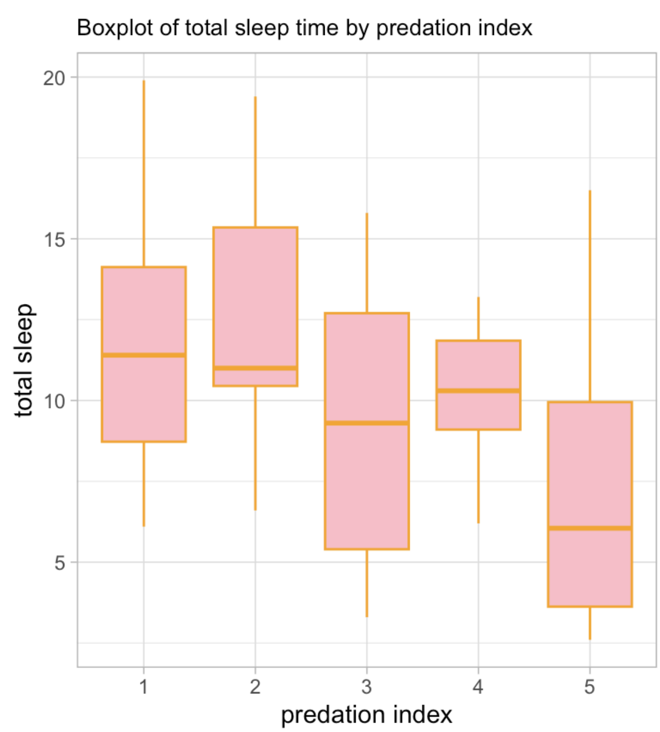


Figure 2: Boxplot of total sleep time in hours per day of mammals for each predation levels

Main analysis:

For the main analysis we used a linear regression test (see figure 3 below).

- Regression equation = -1.24 x (predation_index) + 13.40

The slope of -1.24 indicates that we have a decrease of 1.24 hours of sleep for each increase in predation index. (Intercept is 13.40, but here we have no index 0). The p-value being below our default alpha of 0.05, indicates our result is statistically significant and therefore the predation index does have an effect on the total sleep time. The t value (-3.23) is significantly different from zero, indicating strong evidence we can reject the null hypothesis (H_0= there is no correlation between sleep time and predation). Since the confidence interval for the slope (-2.00 to -0.47) does not include zero this supports the finding that predation has a significant effect on total sleep, as seen in our previous results. Since $F(1, 56)=10.4$, with a p-value <0.05, we can say that the regression model explains a significant portion of the variance, and that the result is statistically significant. R^2 indicates about 15.67% of the variation in total sleep is explained by the predation index, and 14.16% for the adjusted R^2.

We conducted a series of test to check the assumptions were respected:

- Independence of observations: we can confirm from reading the information provided that observations are independent, the sleep time of a sheep doesn't affect the sleep time of a gorilla.
- Linearity: scatter plot with LOESS curve confirms linearity. See figure 4, the linearity is not perfect, but we believe the deviation is not substantial enough to violate the assumption. The residuals vs. fitted value graph shows linearity, figure 5.
- Homoscedasticity with a Breusch-Pagan test (BP(1) = 0.025943, p-value = 0.872) hence our p-value fails to reject the null hypothesis that variance is constant.
- Independence of residuals with a Durbin-Watson test (DW= 1.76, p-value 0.19) fails to reject the null hypothesis that residuals are independent.
- Normality of residuals: see residuals qq plot, figure 6, we see a slight deviation, but we believe it is not substantial enough to violate the assumption.

Discussion:

We can conclude there is evidence to suggest that increased predation index is associated with a reduction in total sleep time for the mammals in your dataset. However, the model explains only a modest amount of the variance in sleep, suggesting other factors may also play a role. Although we've confirmed the independence of observations, it should be noted that animals in a laboratory might not behave in the exact same manner as in the wild and the data should be compared to observations made in the wild.

Visualization:

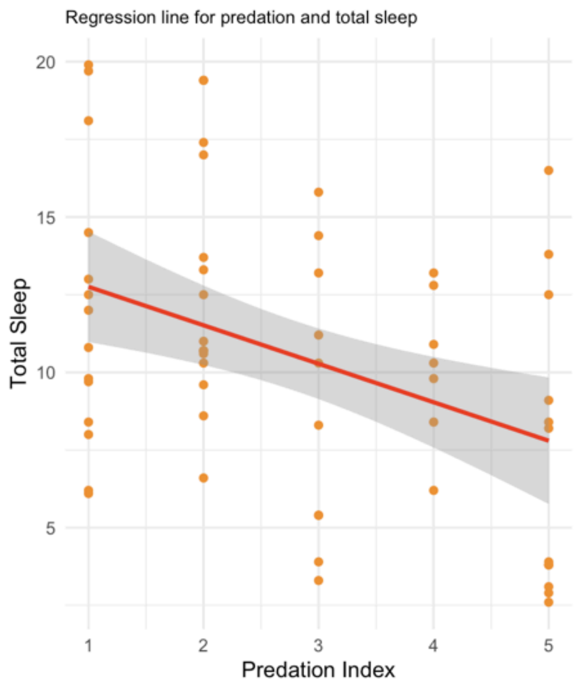


Fig. 3: Regression line showing negative correlation between sleep duration and predation levels. Shaded area shows the confidence interval.

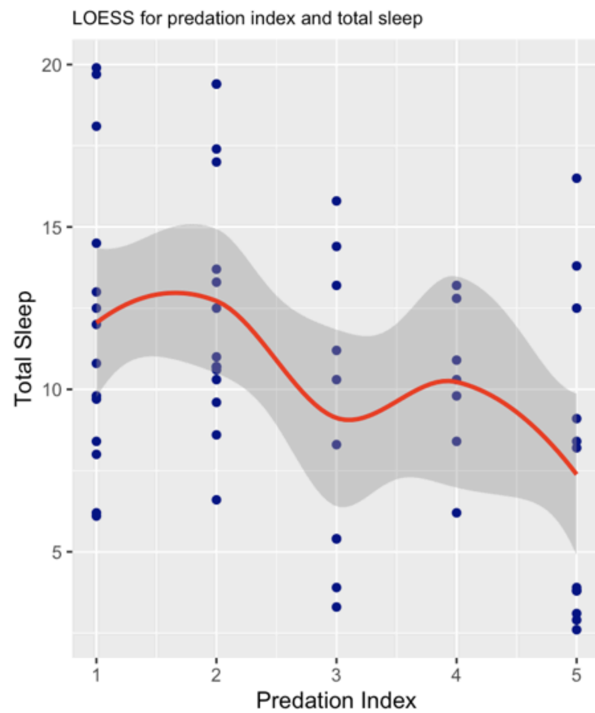


Fig.4: LOESS curve showing the relationship between predation levels(x-axis) and sleep duration (y- axis) with shaded confidence intervals. The trend highlights a decrease of sleep duration as predation levels increase.

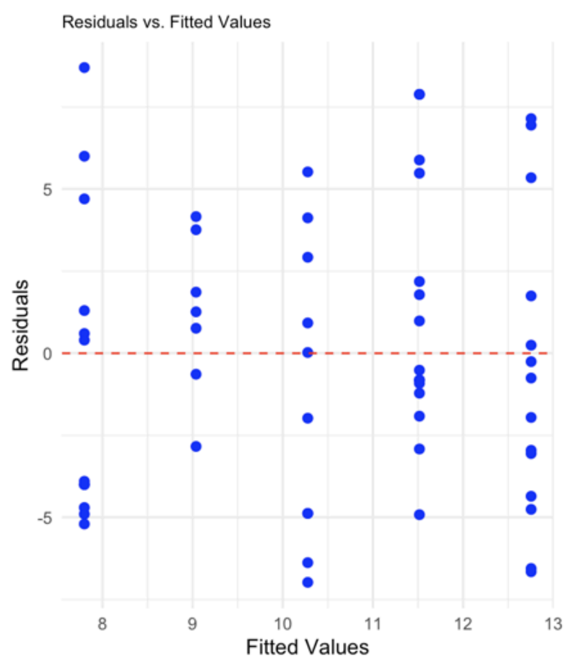


Fig.5: Residuals vs. fitted values shows the residuals are randomly scattered around 0, indicating the model assumptions of linearity and homoscedasticity are met.

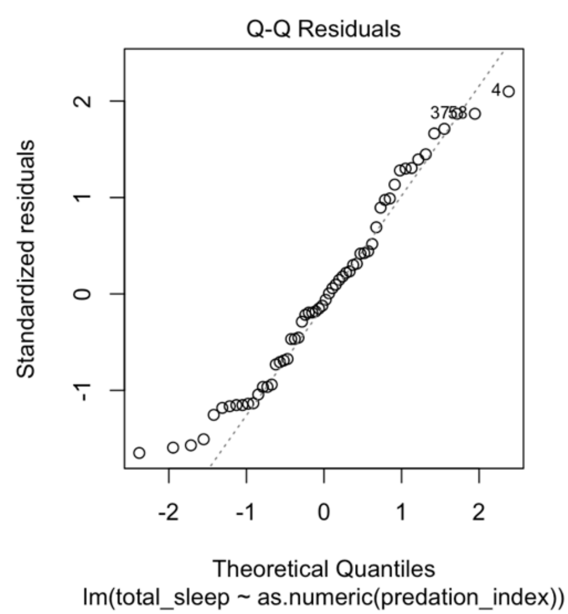


Fig. 6: QQ plot of residuals from the linear regression model showing the slight deviation of residuals from the normal distribution.

Analysis 4: Gestation time vs body weight and sleep time

Rationale and Hypothesis:

Hypothesis: Gestation time positively correlates with sleep time and body weight. Longer gestation time will correlate with longer sleep time, and time for each will increase as an animal's weight increases. Carrying a developing fetus takes lots of energy and therefore more sleep is required for the mother for the duration of the gestation period.

Data preparation:

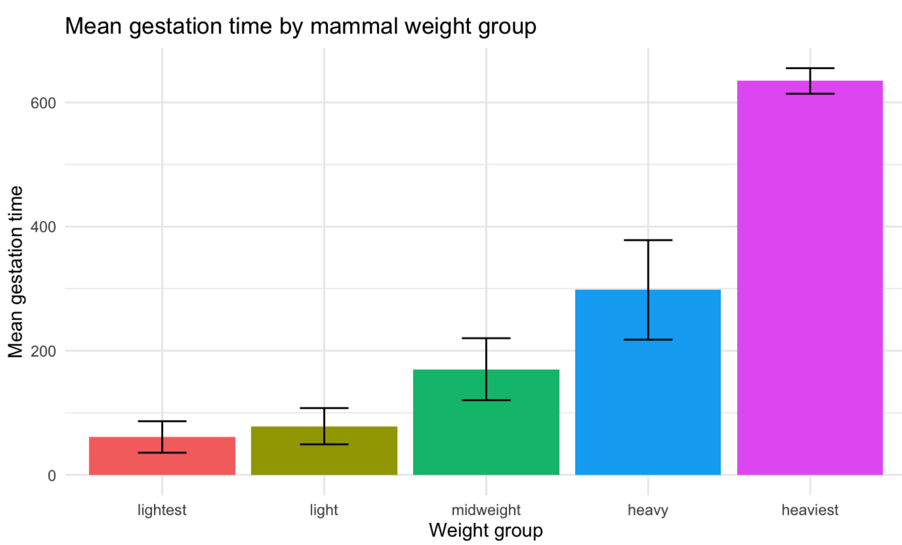
- Added a new column and divisions for grouping the animals by body weight: weight_group

Exploratory Analysis:

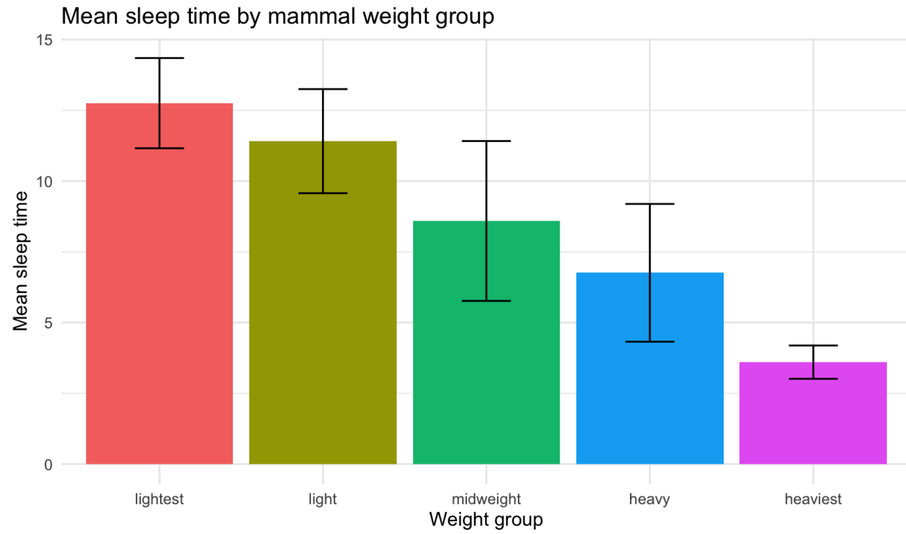
With these basic descriptive statistics, we can see a general trend that mean body weight and gestation time increase as weight group increases, while mean total sleep time decreases:

Variables	lightest (N=19)	light (N=21)	midweight (N=11)	heavy (N=9)	heaviest (N=2)	Overall (N=62)
gestation.time						
Mean (SD)	61.0 (56.4)	78.4 (68.2)	170 (84.4)	298 (123)	635 (14.8)	142 (147)
Median [Min, Max]	42.0 [16.0, 225]	56.0 [12.0, 225]	161 [33.0, 310]	336 [100, 440]	635 [624, 645]	79.0 [12.0, 645]
Missing	2 (10.5%)	1 (4.8%)	1 (9.1%)	0 (0%)	0 (0%)	4 (6.5%)
body.weight						
Mean (SD)	0.266 (0.292)	2.80 (1.50)	40.8 (24.3)	290 (167)	4600 (2900)	199 (899)
Median [Min, Max]	0.120 [0.00500, 0.900]	3.00 [0.920, 6.80]	36.3 [10.0, 85.0]	207 [100, 529]	4600 [2550, 6650]	3.34 [0.00500, 6650]
total.sleep						
Mean (SD)	12.8 (3.55)	11.4 (4.30)	8.59 (4.78)	6.76 (3.73)	3.60 (0.424)	10.5 (4.61)
Median [Min, Max]	12.8 [6.60, 19.9]	10.7 [5.40, 19.4]	8.85 [2.60, 18.1]	6.20 [2.90, 12.0]	3.60 [3.30, 3.90]	10.5 [2.60, 19.9]
Missing	0 (0%)	1 (4.8%)	1 (9.1%)	2 (22.2%)	0 (0%)	4 (6.5%)

The following bar graphs provide an clear visualization of these relationships:



Bar graph demonstrating that gestation time increases as mammal body weight increases



Bar graph demonstrating that sleep time decreases as mammal body weight increases

Main Analysis:

We used correlation analysis to analyze the relationship between sleep time, gestation time and body weight.

Using qq plots, we found that the distribution of the gestation time, body weight, and sleep time data were not normally distributed (see figures 7.1, 7.2, 7.3 below), so we used a Spearman's rank correlation instead of Pearson's. The relationships are monotonic (see figures 8.1, 8.2 below). The results from the correlation tests are as follows:

- Gestation time and sleep time in mammals are significantly, moderately, and negatively correlated (rho = -0.65, $p < 0.001$).
- Body weight and sleep time in mammals are significantly, moderately, and negatively correlated (rho = -0.60, $p < 0.001$).

We also performed a partial correlation test with Spearman's method. After accounting for body weight's impact on a mammal's sleep time, the partial correlation between gestation time and sleep time is negative, weak, and significant (rho = -.26, $p = 0.01$). After accounting for gestation time's impact on a mammal's sleep time, the partial correlation between body weight and sleep time is negative, weak, and not significant (rho = -.18, $p = 0.07$).

Visualizations:

QQ plots:

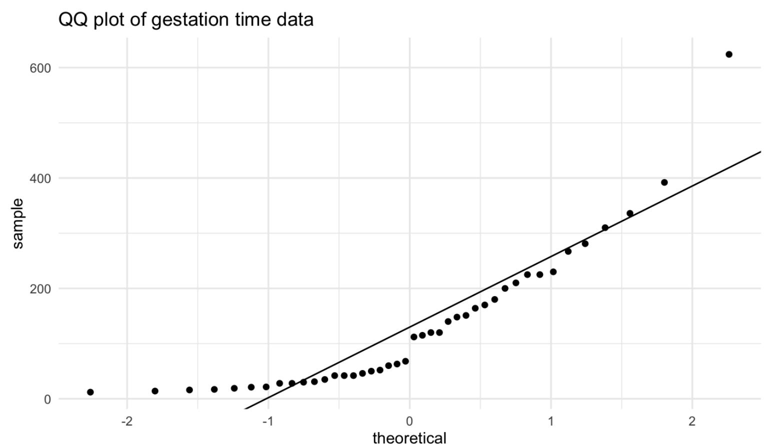


Figure 7.1

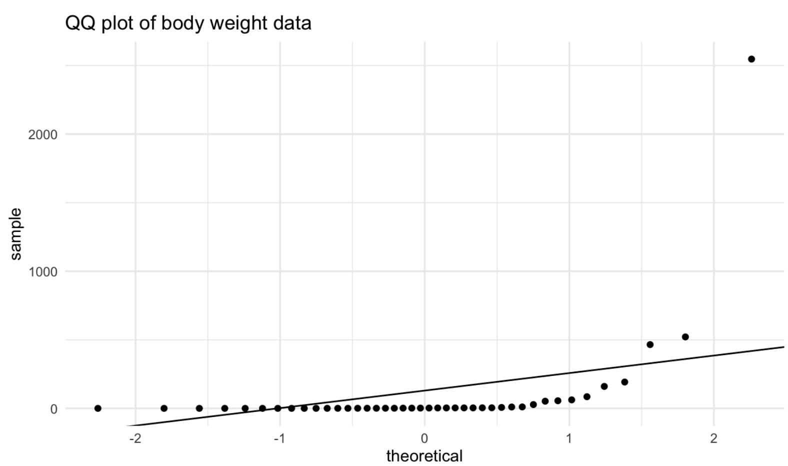


Figure 7.2

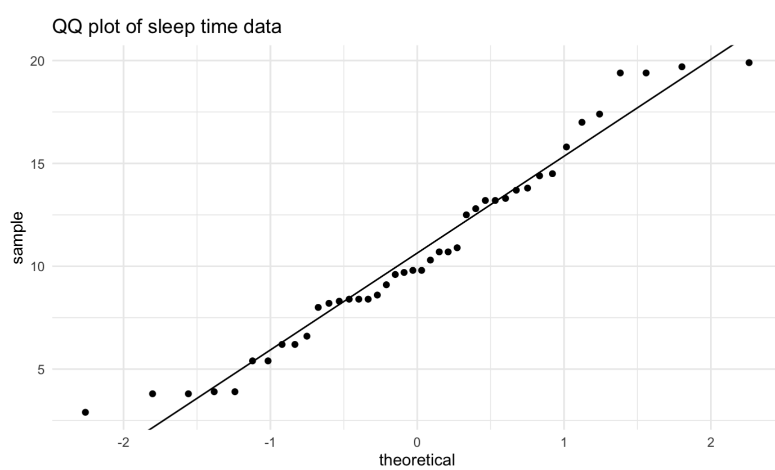


Figure 7.3

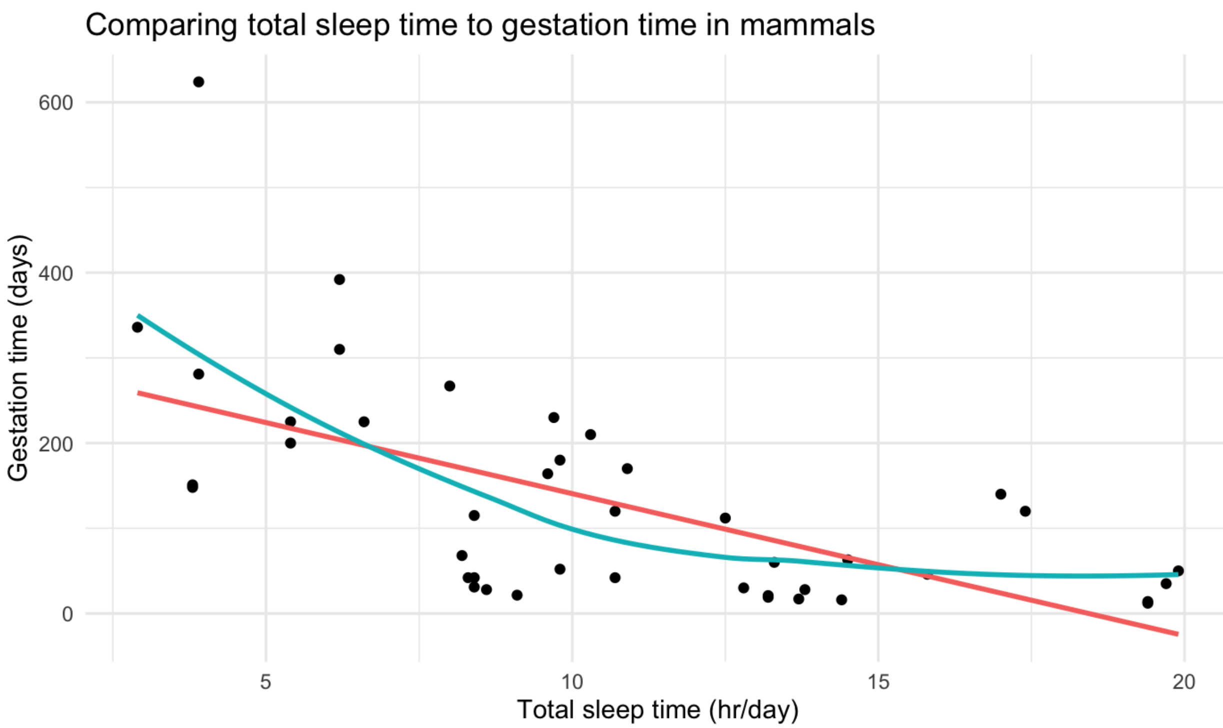


Figure 8.1: Scatter plot demonstrating relationship between sleep time and gestation time

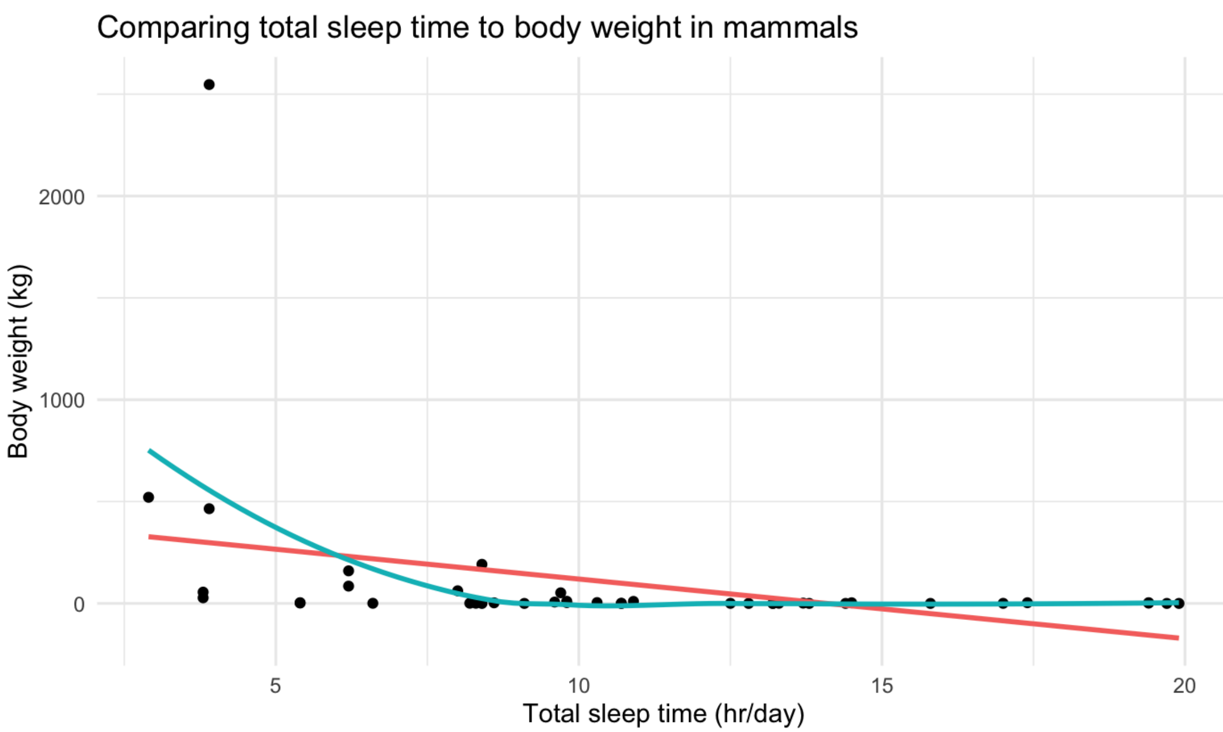


Figure 8.2: Scatter plot demonstrating relationship between sleep time and body weight

Discussion:

We discovered that the data does not support our original hypothesis: there is actually a negative correlation between the gestation time and sleep time among mammals rather than a positive one, as well as a negative correlation between body weight and sleep time. We also found that body weight positively correlates to gestation time. We can summarize these findings to say that as a mammal's size increases, their gestation time also increases, however as a mammal's size and gestation time increase, then the amount of time they sleep per day decreases. Perhaps this reveals that the bigger an animal is, the less time they need to sleep per day so they must carry their growing offspring for a longer period of time. There are additional factors that impact the amount of time that a mammal sleeps per day that were not taken into account in this analysis; it would be interesting to include additional factors in the analysis to see how they might impact the results.