Stat216 FINAL PROJECT

Scarlett QI

2023-04-14

```
library(readr)
Sleep_data <- read_csv("C:/Users/scarl/Desktop/STAT216/Sleep data.csv")
```

```
## Rows: 62 Columns: 8
## -- Column specification ---------------------------------------------------------
## Delimiter: ","
## chr (3): max_life_span, gestation_time, total_sleep
## dbl (5): body_weight, brain_weight, predation_index, sleep_exposure_index, d...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
View(Sleep_data)
nrow(Sleep_data)
```

```
## [1] 62
```

```
sleep2 <- Sleep_data[!(Sleep_data$body_weight=="?" | Sleep_data$brain_weight=="?" | Sleep_data$max_life_
sleep2
```

```
## # A tibble: 51 x 8
##    body_weight brain_weight max_life_s~1 gesta~2 preda~3 sleep~4 dange~5 total~6
##          <dbl>        <dbl> <chr>        <chr>     <dbl>   <dbl>   <dbl> <chr>
## 1        6654          5712 38.6         645           3       5       3 3.3
## 2           1           6.6 4.5          42            3       1       3 8.3
## 3           3.38        44.5 14          60            1       1       1 12.5
## 4        2547          4603 69           624           3       5       4 3.9
## 5          10.6         180. 27          180           4       4       4 9.8
## 6           0.023         0.3 19         35            1       1       1 19.7
## 7         160           169 30.4         392           4       5       4 6.2
## 8           3.3          25.6 28         63            1       2       1 14.5
## 9          52.2         440 50           230           1       1       1 9.7
## 10          0.425         6.4 7          112           5       4       4 12.5
## # ... with 41 more rows, and abbreviated variable names 1: max_life_span,
## #   2: gestation_time, 3: predation_index, 4: sleep_exposure_index,
## #   5: danger_index, 6: total_sleep
```

```
nrow(sleep2)
```

```
## [1] 51
```

```
sleep2$total_sleep <- as.numeric(sleep2$total_sleep)
sleep2$sleep_exposure_index <- as.factor(sleep2$sleep_exposure_index)
```

Part 1: Background information

The data set we want to explore collected brain weight, body weight, life span, gestation time, time sleeping, predation, and danger indices for 62 mammals. In our study, we mainly focus on brain weight, maximum life span, predation index, sleep exposure index, and total sleep.

Brain weight is the mammals' brain weight in g, ranging from 0.14g to 5712g, with mean value 317.5. Boxplot and histogram show that this data skews to the right and have extreme outliers. However, it is normal to have outliers in this case since mammals' brain size and weight differ a lot.

```
par(mfcol=c(1,2))
summary(sleep2$brain_weight)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.14    4.50   21.00  317.50  172.00 5712.00
```

```
boxplot(sleep2$brain_weight)
m <- mean(sleep2$brain_weight)
points(x=1,y=m,col="blue")
hist(sleep2$brain_weight)
```
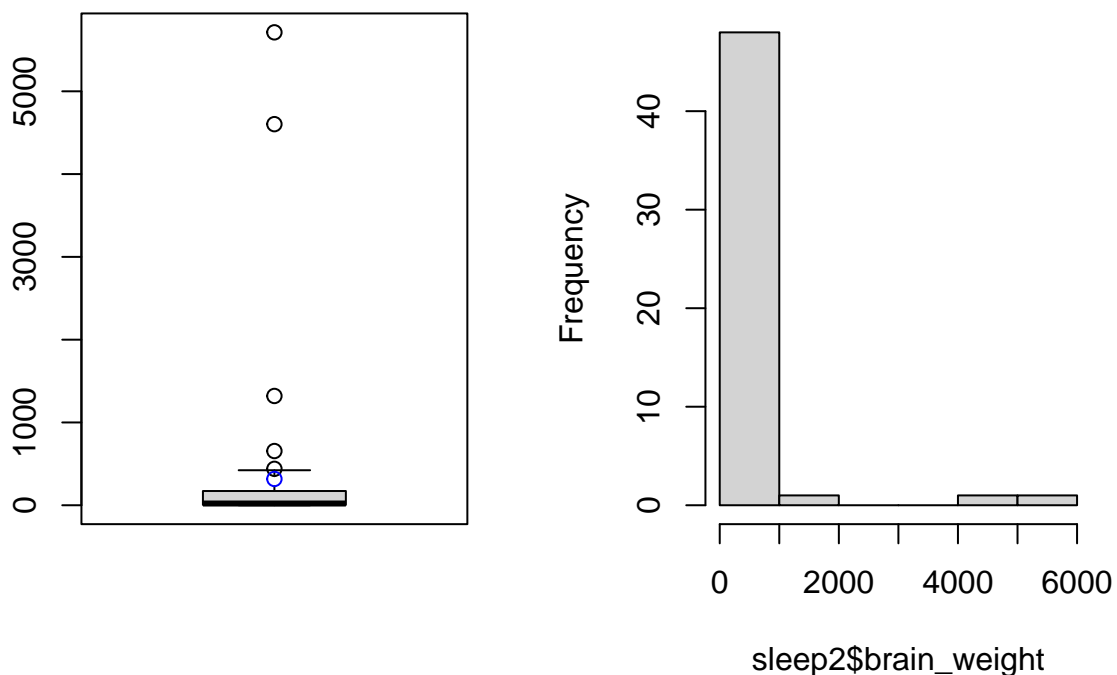
Maximum life span is the length of time for which mammals live in years. This variable ranges from 2 years to 100 years, and the average maximum life span is 20.15 years. Both boxplot and histogram are right skewed and extreme outliers exist, indicating that this data is not normal. However, whether this distribution is normal or not is not important since we use max life span as a variable in Q|QQC case, where normality assumptions is checked via the histogram of total sleep time and qq plot of residuals.

```
par(mfcol=c(1,2))
summary(sleep2$max_life_span)
```

```
##     Length     Class      Mode
##         51 character character
```

```
sleep2$max_life_span<- as.numeric(sleep2$max_life_span)
boxplot(sleep2$max_life_span)
m <- mean(sleep2$max_life_span)
points(x=1,y=m,col="blue")
hist(sleep2$max_life_span)
```
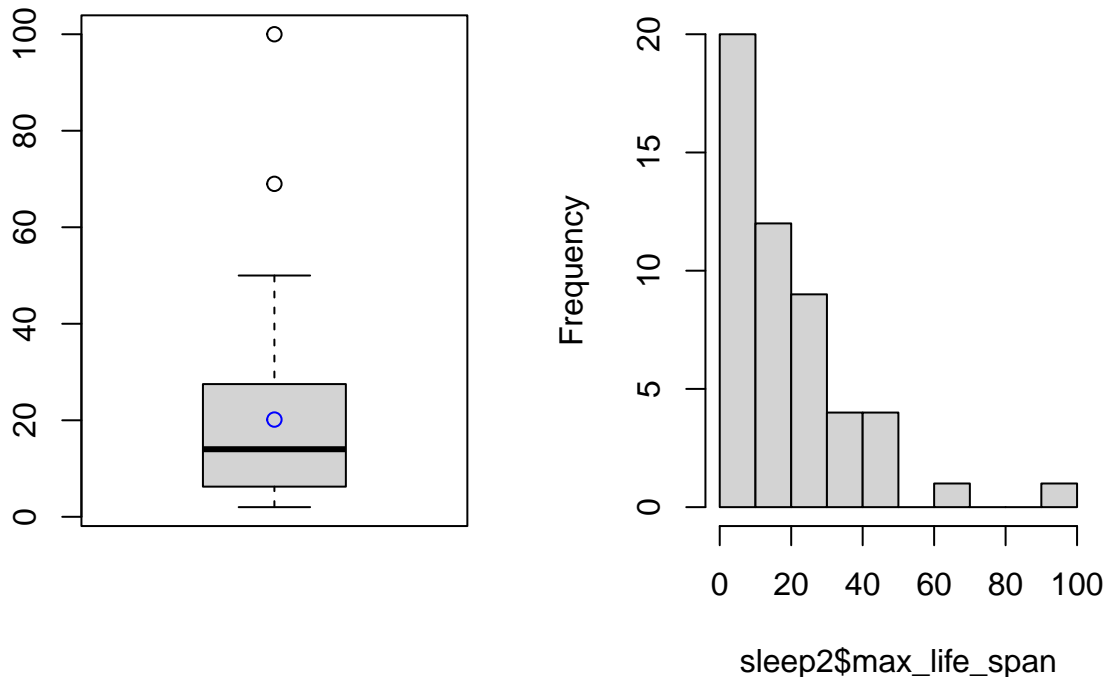

Predation index means the probability that one organism preyed and killed by another. This index is a categorical variable scores from 1 to 5, 1 represents least likely to be preyed upon and 5 means this mammal is most likely to be preyed upon.

```
par(mfcol=c(1,2))
summary(sleep2$predation_index)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000   2.000   3.000   2.843   4.000   5.000
```

```
sleep2$predation_index<- as.numeric(sleep2$predation_index)
hist(sleep2$predation_index)
```

Histogram of sleep2$predation_in


sleep2$predation_index
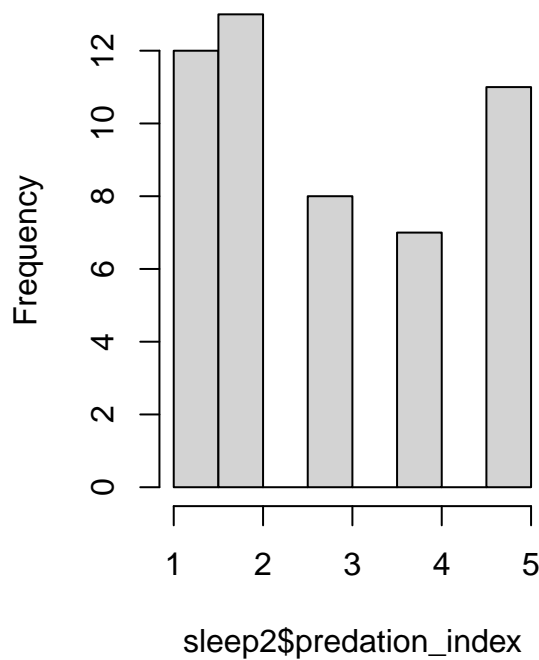
As another categorical variable, sleep exposure index also scores from 1 to 5, 1 means this mammal is least exposed and sleeps in a well-protected den. 5 means it's most exposed.

```
par(mfcol=c(1,2))
summary(sleep2$sleep_exposure_index)
```

```
##  1  2  3  4  5
## 21 11  4  5 10
```

```
sleep2$sleep_exposure_index<- as.numeric(sleep2$sleep_exposure_index)
hist(sleep2$sleep_exposure_index)
```

stogram of sleep2$sleep_exposure
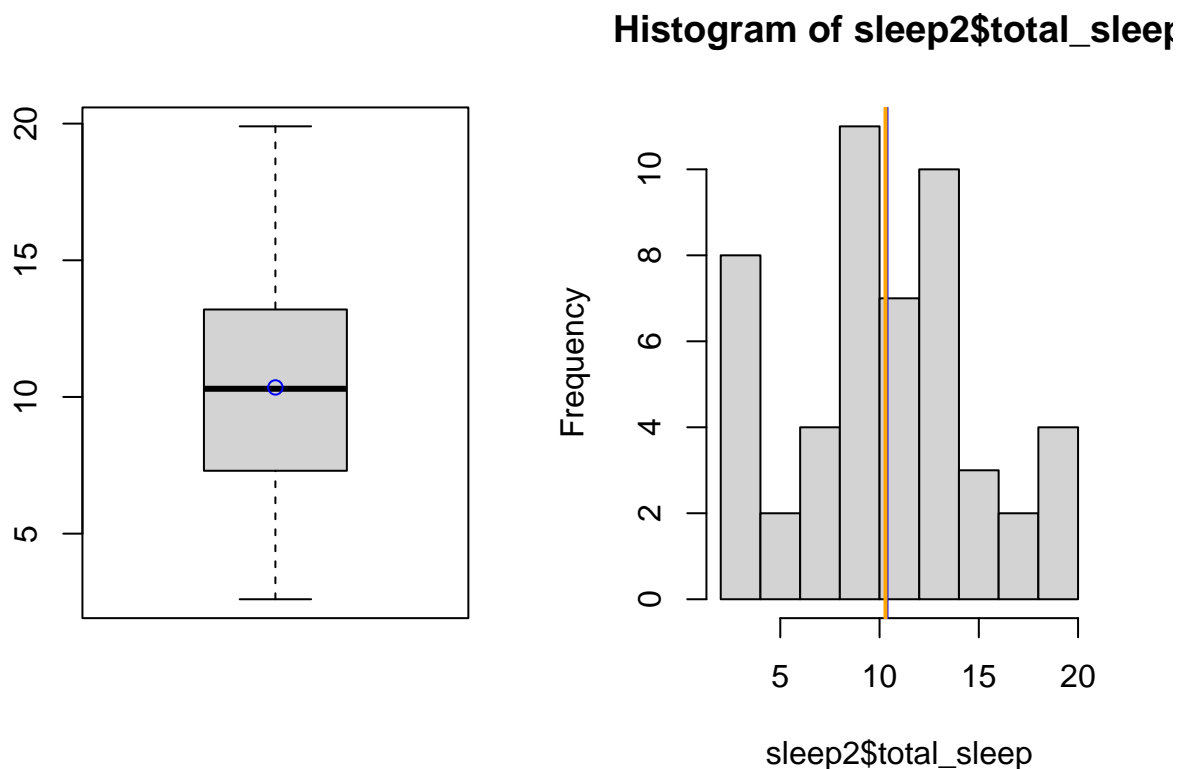

sleep2$sleep_exposure_index
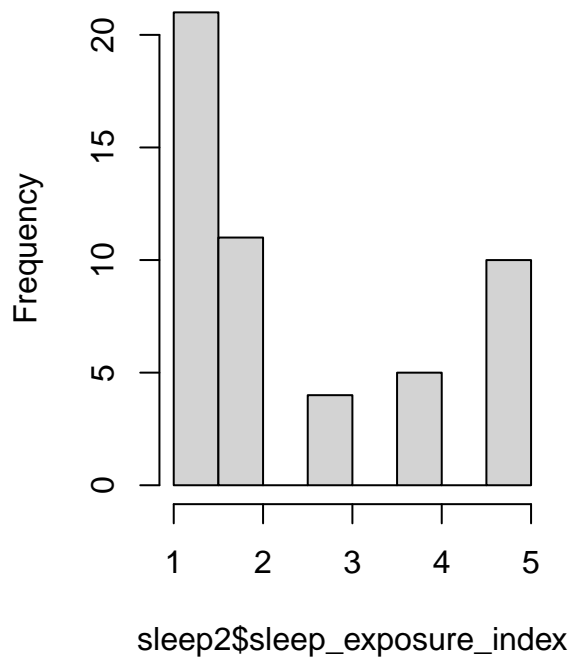
Total sleep represents the mammals' total sleep in hours. Total sleep ranges from 2.6 hours to 19.9 hours, with mean value 10.35 hours and median 10.3 hours. The mean and median are pretty close to each other. The boxplot and the histgram also show that this variable has a normal distribution, without any outliers.

```
par(mfcol=c(1,2))
sleep2$total_sleep<- as.numeric(sleep2$total_sleep)
summary(sleep2$total_sleep)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.60    7.30   10.30   10.35   13.20   19.90
```

```
boxplot(sleep2$total_sleep)
m <- mean(sleep2$total_sleep)
points(x=1,y=m,col="blue")
hist(sleep2$total_sleep)
abline(v = mean(sleep2$total_sleep), col="blue",lwd=2)
abline(v = median(sleep2$total_sleep), col="orange",lwd=2)
```

Histogram of sleep2$total_sleep

sleep2$total_sleep

Part 2: Analysis of Case CQ

Big hypothesis

Sleep exposure significantly correlate with total sleep time.

Rationale

Sleep exposure influences animals' sense of security. Essentially, we are exploring if the sense of security would correlate with total sleep time. People always have the common sense that there will be some correlation between total sleep time and the sense of security. Therefore, our analysis will give numerical evidence inference on this topic.

Project Plan with R code and results analysis

[STEP 1] : Checking assumptions for doing an ANOVA

Assumption 1 — independent among groups Obviously populations with different sleep exposure index are independent from each other and there are no pairs or correlation.

Assumption 2 — independent among subjects Observation within groups are also independent, since the measurements for each mammal are in no way influenced by the measurements of other mammals.

Assumption 3 — homogeneity We use Levene's Test to check homogennity. The p-value of this test is 0.09, which is larger α, thus we fail to reject H0 and can conclude that these five groups with different sleep exposure index have equal variances. Therefore homogeneity holds.

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.2.3
```

```
## Loading required package: carData
```

```
sleep2$sleep_exposure_index<- as.factor(sleep2$sleep_exposure_index)
leveneTest(total_sleep~sleep_exposure_index,data=sleep2, center = median)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value  Pr(>F)
## group  4  2.1298 0.09226 .
##       46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Assumption 4 — normality We have an unbalanced design with 21, 11, 4, 5 and 10 observations in each exposure group with different index. Moreover, the sample size of each group is less than 30, being small sample sizes so that we can not apply CLT. Looking at boxplot of each sample, mean is approximately equal to median. Further looking at numerical test, the p-values of shapiro tests with exposure index 1 to 4 are all smaller than 0.05, indicating normality. However, the p-value of shapiro test with exposure index 5 being 0.00526 is smaller than 0.05, violating the normality assumption. All in all, the normality assumption does not hold, we decide to use Kruskal-Wallis Test instead of traditional ANOVA.

```
table(sleep2$sleep_exposure_index)
```

```
##
##  1  2  3  4  5
## 21 11  4  5 10
```

```
par(mfrow=c(1,2))
```

```
boxplot(total_sleep ~ sleep_exposure_index,data=sleep2, main="Boxplot of total sleep time and sleep exp
(group.means <- tapply(sleep2$total_sleep, sleep2$sleep_exposure_index, mean))
```

```
##        1        2        3        4        5
## 13.14762 11.08182  8.57500 10.72000  4.19000
```

```
points(1:5, group.means)
```

```
tapply(sleep2$total_sleep,sleep2$sleep_exposure_index, shapiro.test)
```
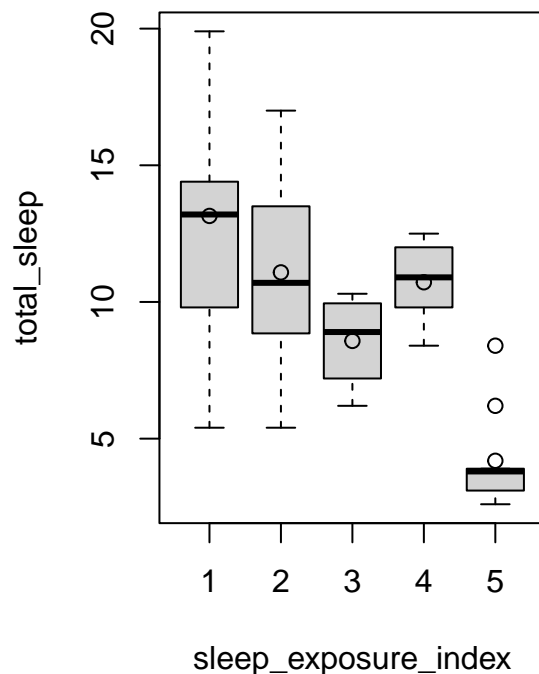
```
## $'1'
##
##  Shapiro-Wilk normality test
##
```

```
## data:  X[[i]]
## W = 0.93652, p-value = 0.1859
##
##
## $`2`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.96966, p-value = 0.8831
##
##
## $`3`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.94892, p-value = 0.7094
##
##
## $`4`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.95786, p-value = 0.793
##
##
## $`5`
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.76383, p-value = 0.00526
```

ot of total sleep time and sleep expo

sleep_exposure_index

[STEP 2] : doing Kruskal-Wallis Test The p-value of Kruskal-Wallis test is 3.414e-05, which is pretty small and less than 0.05. Thus we have enough evidence to reject H0 and therefore we conclude that at least one group mean differs from another.

```
kruskal.test(total_sleep ~ sleep_exposure_index, data=sleep2)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  total_sleep by sleep_exposure_index
## Kruskal-Wallis chi-squared = 25.837, df = 4, p-value = 3.414e-05
```

[STEP 3] : conducting multiple comparisons Since the p-value of Kruskal-Wallis Test is significant. We delve deeper to see where the difference in total sleep time lie. From Kurskal multiple comparison, we can see that the mean of total sleep time with index 1 and 5 and the mean of total sleep time with index 2 and 5 are significantly different, which is actually indicating that the level of exposure do associated with total sleep time.

```
library(pgirmess)
```

```
## Warning: package 'pgirmess' was built under R version 4.2.3
```

```
kruskalmc(total_sleep ~ sleep_exposure_index, data=sleep2)
```

```
## Multiple comparison test after Kruskal-Wallis
## alpha: 0.05
## Comparisons
##        obs.dif critical.dif stat.signif
## 1-2  5.759740    15.53148       FALSE
## 1-3 16.089286    22.76534       FALSE
## 1-4  7.014286    20.76518       FALSE
## 1-5 28.164286    16.03301        TRUE
## 2-3 10.329545    24.36481       FALSE
## 2-4  1.254545    22.50725       FALSE
## 2-5 22.404545    18.23296        TRUE
## 3-4  9.075000    27.99304       FALSE
## 3-5 12.075000    24.68754       FALSE
## 4-5 21.150000    22.85622       FALSE
```

[STEP 4] : calculating correlation between sleep exposure index and total sleep time We divide mammals into "low exposure" and "high exposure" according to their sleep exposure index, and calculate the bi-serial correlation between total sleep time and exposure level. The bi-serial correlation is -0.7474052, indicating a negative and moderately strong relationship. The p-value of correlation test is 2.947115e-05, which is smaller than α. Thus we have enough evidence to reject H0 and conclude that this Biserial correlation is significant, and can happen at the population level based on our sample.

```
sleep2$sleep_exposure_index <- as.numeric(sleep2$sleep_exposure_index)
sleep2$total_sleep <- as.numeric(sleep2$total_sleep)

sleep2$sleep_exposure_index <- cut(sleep2$sleep_exposure_index, c(-Inf, 2, Inf), c("low exposure", "high
sleep2$sleep_exposure_index_new <- ifelse(sleep2$sleep_exposure_index == "low exposure", 0, 1)

library(polycor)
```

```
## Warning: package 'polycor' was built under R version 4.2.3
```

```
polyserial(sleep2$total_sleep, sleep2$sleep_exposure_index)
```

```
## [1] -0.7474052
```

```
rb <- polyserial(sleep2$total_sleep, sleep2$sleep_exposure_index_new)
n <- nrow(sleep2)
tab1 <- prop.table(table(sleep2$sleep_exposure_index_new))
p1 <- max(tab1)
p2 <- min(tab1)
y <- dnorm(qnorm(p1))
sd <- sqrt(p1*p2)/(y*sqrt(n))
z <- rb/sd
2*pnorm(-abs(z))
```

```
## [1] 2.947115e-05
```

Part 3: Analysis of Case Q|QQC

Big hypothesis

we can predict the total sleep time based on brain weight, max life span, and predation index.

Ratinale

Brain weight, life span and predation index are three factors that people don't usually think about when discussing sleep. We are interested that if those seemingly "not directly related" factor do associated with sleep time, or we can use them to predict sleep time.

Project plan with R code and results analysis

[STEP 1] : fitting a multiple regression model

AIC for forward elimination and stepwise selection are both 283.14995, whereas AIC is 299.6893 for backward selection. Forward elimination and stepwise selection both reached the same model, and this model had a lower AIC than the model from backward selection. Based on this, the preferred model for total sleep time is based on max_life_span and predation_index.

The model is

$$y_hat = 15.4471 - 0.1113 * x_{max_life_span} - 0.8986 * x_{predation2} - 4.8357 * x_{predation3} - 3.0913 * x_{predation4} - 6.6951 * x_{predation5}$$

```
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 4.2.3
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
sleep2$max_life_span <- as.numeric(sleep2$max_life_span)
sleep2$predation_index <- as.factor(sleep2$predation_index)
lm_full <- lm(total_sleep ~ brain_weight + max_life_span + predation_index, data=sleep2)

ols_step_forward_aic(lm_full, details=FALSE)
```

```
##
##                         Selection Summary
## --------------------------------------------------------------------------
## Variable          AIC        Sum Sq       RSS         R-Sq      Adj. R-Sq
## --------------------------------------------------------------------------
## predation_index   296.119    310.480     784.387     0.28358     0.22128
## max_life_span     283.150    509.993     584.874     0.46580     0.40645
## --------------------------------------------------------------------------
```

```
forward <- lm(total_sleep ~ max_life_span + predation_index, data=sleep2)
summary(forward)
```

```
##
## Call:
## lm(formula = total_sleep ~ max_life_span + predation_index, data = sleep2)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1695 -2.4227 -0.3841  1.9101  7.1238
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      15.4471     1.3559  11.392 7.51e-15 ***
## max_life_span    -0.1113     0.0284  -3.918 0.000301 ***
## predation_index2 -0.8986     1.5247  -0.589 0.558564
## predation_index3 -4.8357     1.6899  -2.862 0.006374 **
## predation_index4 -3.0913     1.7454  -1.771 0.083306 .
## predation_index5 -6.6951     1.5361  -4.358 7.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.605 on 45 degrees of freedom
## Multiple R-squared:  0.4658, Adjusted R-squared:  0.4064
## F-statistic: 7.848 on 5 and 45 DF,  p-value: 2.222e-05
```

```
AIC(forward)
```

```
## [1] 283.1499
```

```
ols_step_backward_aic(lm_full, details=FALSE)
```

```
##
##
##                     Backward Elimination Summary
## --------------------------------------------------------------------------
## Variable          AIC         RSS        Sum Sq       R-Sq       Adj. R-Sq
## --------------------------------------------------------------------------
## Full Model      284.561     578.154     516.713     0.47194      0.39993
## brain_weight    283.150     584.874     509.993     0.46580      0.40645
## --------------------------------------------------------------------------
```

```
backward <- lm(total_sleep ~ brain_weight, data=sleep2)
summary(backward)
```

```
##
## Call:
## lm(formula = total_sleep ~ brain_weight, data = sleep2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1200 -2.5296 -0.1777  2.3165  9.0143
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.886168   0.645119  16.875  < 2e-16 ***
## brain_weight -0.001692   0.000610  -2.774  0.00782 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.395 on 49 degrees of freedom
## Multiple R-squared:  0.1357, Adjusted R-squared:  0.1181
## F-statistic: 7.693 on 1 and 49 DF,  p-value: 0.007822
```

```
AIC(backward)
```

```
## [1] 299.6893
```

```
ols_step_both_aic(lm_full, details=FALSE)
```

```
##
##
##                              Stepwise Summary
## -------------------------------------------------------------------------------
## Variable           Method       AIC        RSS       Sum Sq      R-Sq     Adj. R-Sq
## -------------------------------------------------------------------------------
## predation_index    addition    296.119    784.387    310.480    0.28358    0.22128
## max_life_span      addition    283.150    584.874    509.993    0.46580    0.40645
## -------------------------------------------------------------------------------
```

```
both <- lm(total_sleep ~ max_life_span + predation_index, data=sleep2)
summary(both)
```

```
##
## Call:
## lm(formula = total_sleep ~ max_life_span + predation_index, data = sleep2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1695 -2.4227 -0.3841  1.9101  7.1238
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       15.4471     1.3559  11.392 7.51e-15 ***
## max_life_span     -0.1113     0.0284  -3.918 0.000301 ***
## predation_index2  -0.8986     1.5247  -0.589 0.558564
## predation_index3  -4.8357     1.6899  -2.862 0.006374 **
## predation_index4  -3.0913     1.7454  -1.771 0.083306 .
## predation_index5  -6.6951     1.5361  -4.358 7.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.605 on 45 degrees of freedom
## Multiple R-squared:  0.4658, Adjusted R-squared:  0.4064
## F-statistic: 7.848 on 5 and 45 DF,  p-value: 2.222e-05
```

```
AIC(both)
```

```
## [1] 283.1499
```

[STEP 2] : check assumptions

Assumption 1 — Independent errors From he index plot of the residuals, we see random scatter about the y = 0 line, and there are almost the same observations above and below the y=0 line. We also use Durbin Watson Test to check Autocorrelation. The p-value for the Durbin Watson is 0.056, which is larger than 0.05. Thus p > 0.05 and we conclude that the errors are not auto-correlated. Thus independence assumption is not violated.
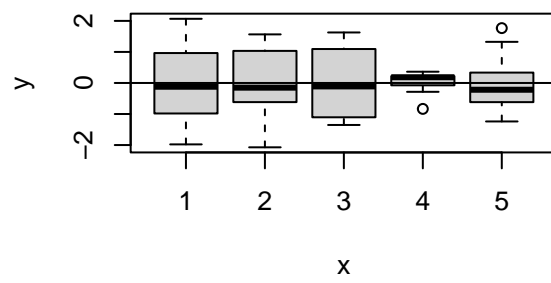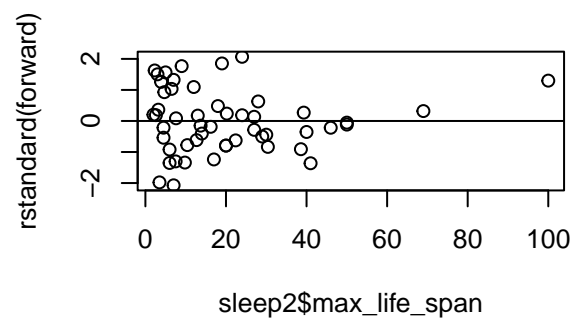
```
plot(rstandard(forward), type="b")
abline(h=0)
```
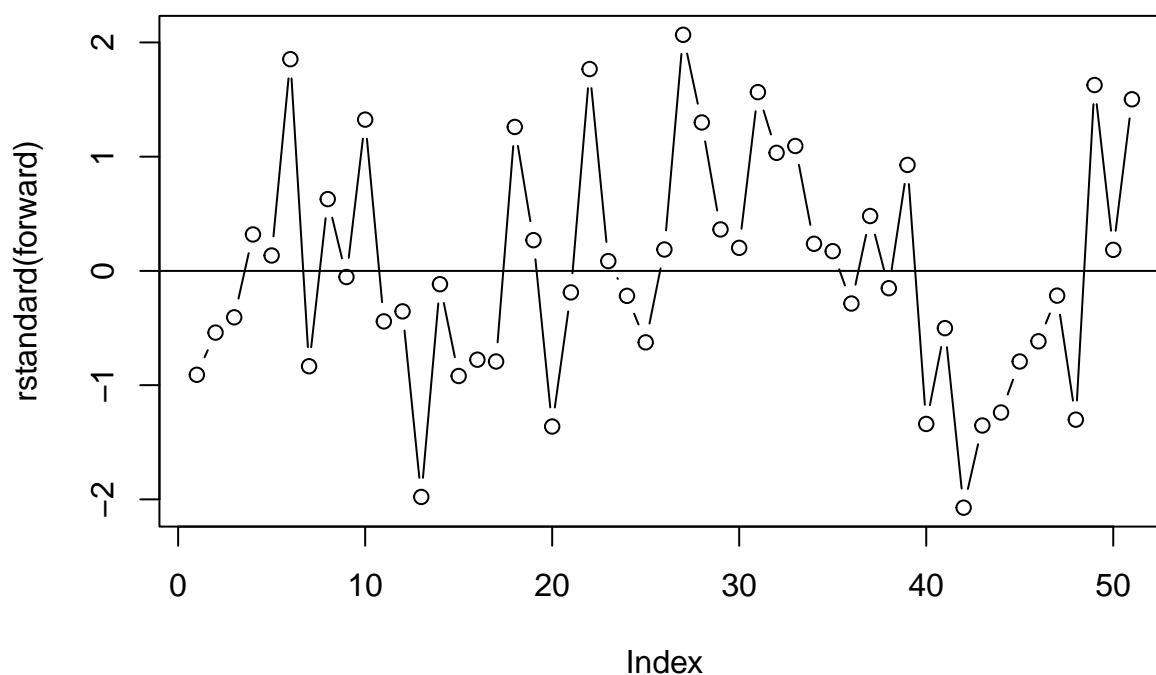

Assumption 2 — linearity we check nonlinearity by plotting the standard residuals versus max life span and predation index. Since predation index is a categorical variable, the plot for predation index isn't too meaningful. In another plot regarding max life span, we generally see random scatter. There are almost same numbers of observations above and below the y=0 line and there's no serious violations of linearity. According to the resdual fitted plot, we notice that there is no obvious pattern and the red line is approximately horizontal about y = 0, suggesting that we can assume linear relationship between the predictors and the outcome variables.

```
par(mfrow=c(2,2))
plot(sleep2$max_life_span,rstandard(forward))
abline(h=0)
plot(sleep2$predation_index,rstandard(forward))
abline(h=0)

par(mfrow=c(1,1))
```

14

```
plot(forward, 1)
```

Residuals vs Fitted

Fitted values
lm(total_sleep ~ max_life_span + predation_index)
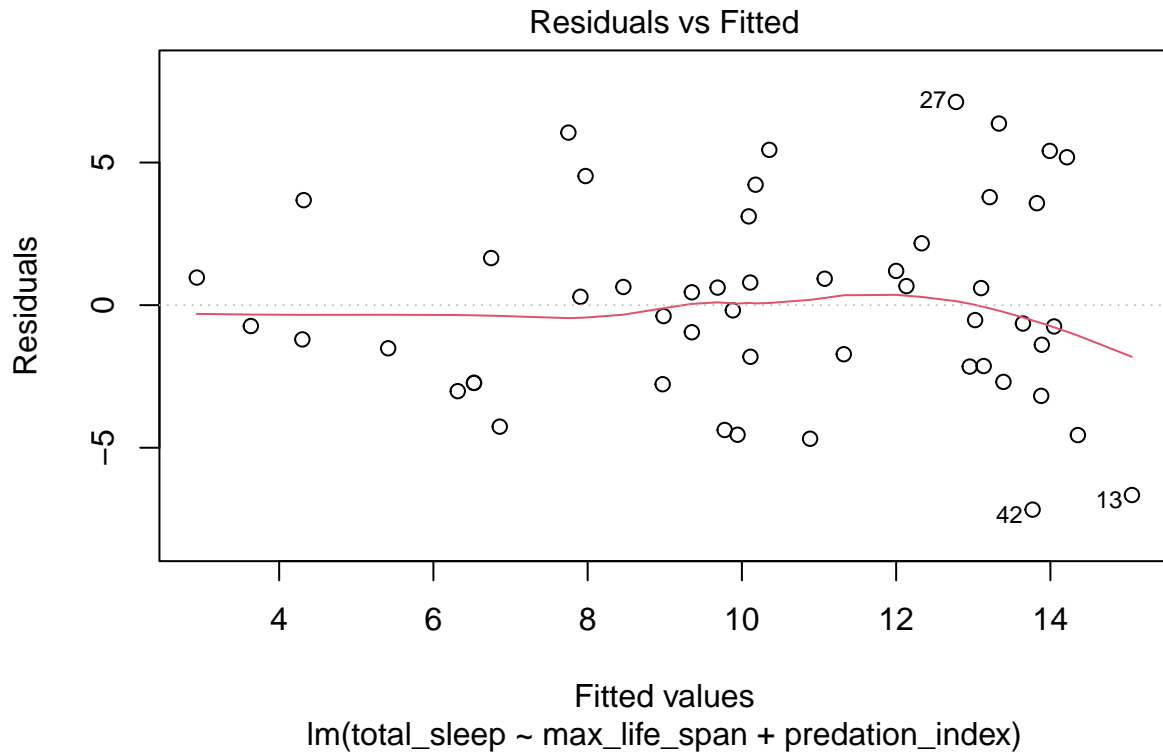
assumption 3 — homoscedacity In examination of the residual(or standardized residuals) versus fitted plots, we see the red line should be approximately horizontal at zero. Moreover, the p-value of Breusch-Pagan test is 0.05088, which is greater than 0.05. Overall, the constant variance assumption seems reasonable.

assumption 4 —normality In qq plot, almost all points are on the diagonal line. The distribution of total sleep time approximately follows normal distribution. Moreover, the p-value of shapiro test being 0.4554 is greater than 0.05. All in allm normality assumption is met.

```
plot(forward, 2)
```

Normal Q–Q

Theoretical Quantiles
lm(total_sleep ~ max_life_span + predation_index)
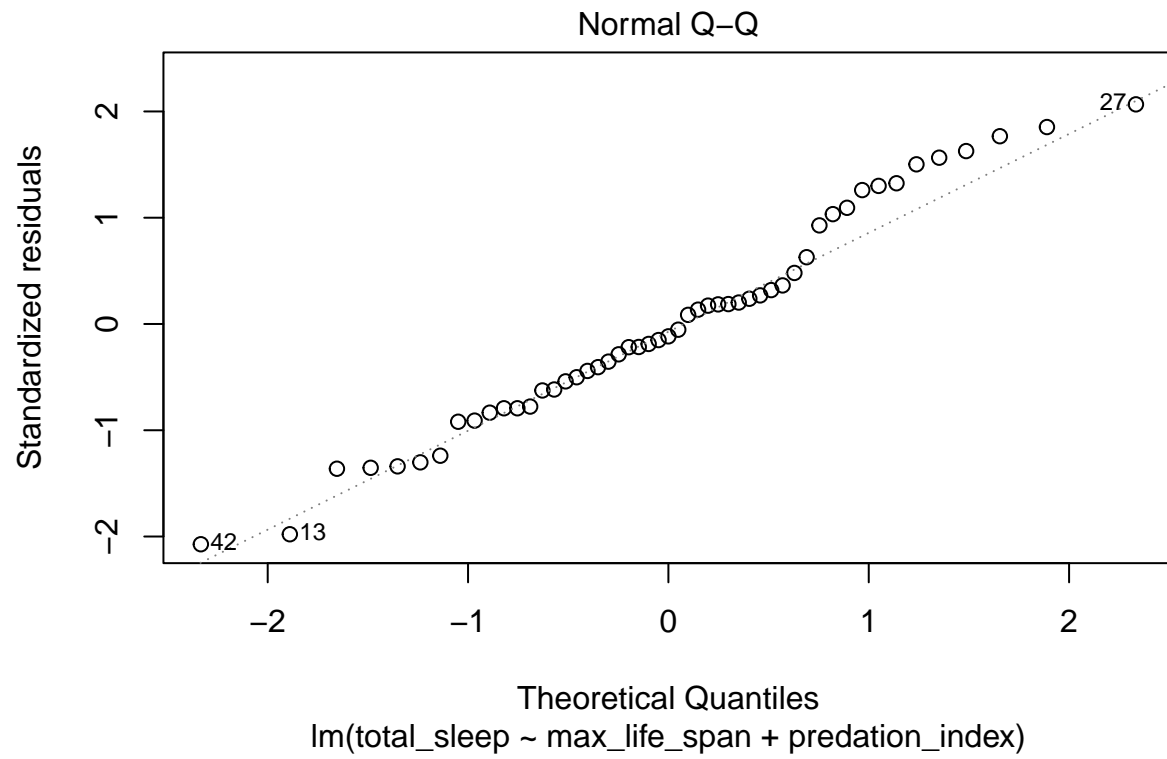
```
sleep2$total_sleep <- as.numeric(sleep2$total_sleep)
hist(sleep2$total_sleep)
```

Histogram of sleep2$total_sleep
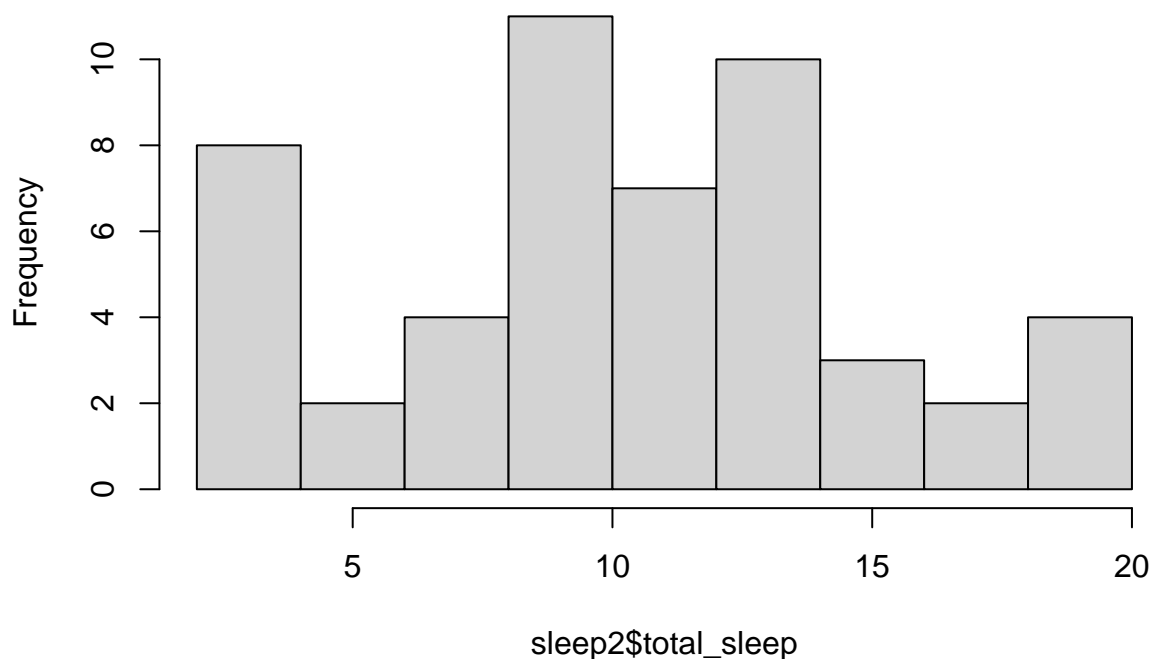

```
resid1 <- resid(forward)
shapiro.test(resid1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid1
## W = 0.97794, p-value = 0.4554
```

Assumption 5 — multicollinearity Since we don't want collinearity among the predictor variables, we use pairs of scatter plots. Since predation index is a functionally categorical data, and is not continuous, the scatter plots are not meaningful to explore. Instead, we focus on VIF value. We see that the VIF values are all below 10, thus we can conclude that max life span and predation index are relatively unrelated.

```
par(mfrow=c(2,3))

pairs(sleep2[,c("max_life_span", "predation_index")])
```
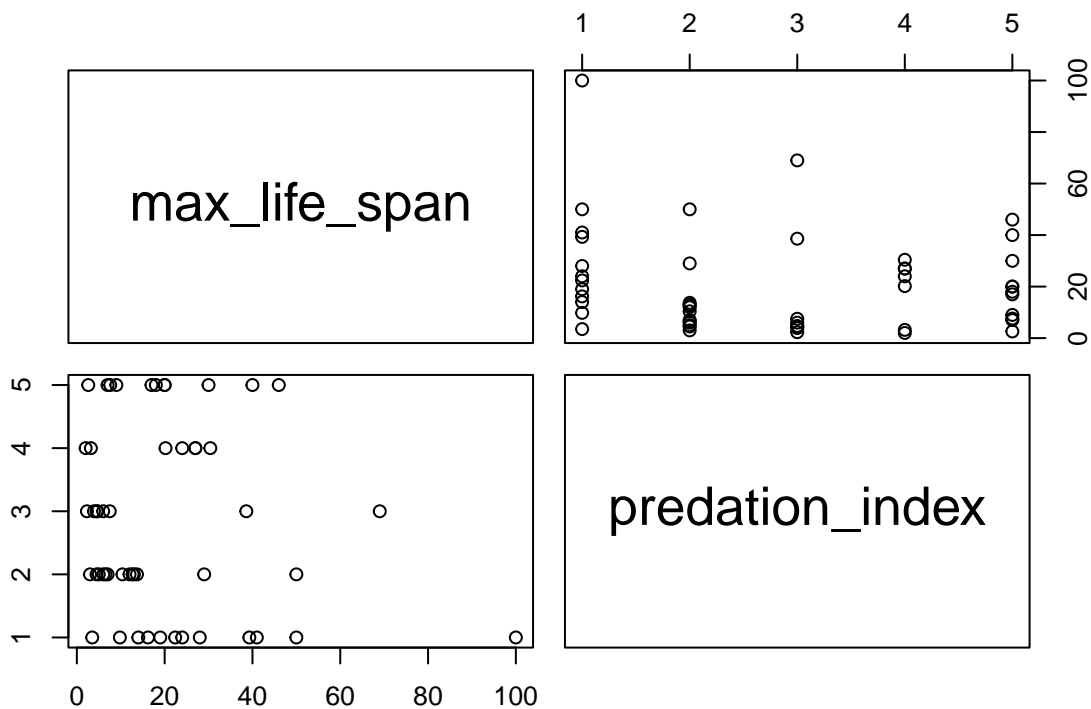
```
##cor(sleep2[,c("max_life_span", "predation_index")])
library(car)
vif(forward)
```

```
##                   GVIF Df GVIF^(1/(2*Df))
## max_life_span   1.12461  1        1.060476
## predation_index 1.12461  4        1.014788
```

Assumption 6 — y variable to be quantitative Assumption holds since the outcome variable: total sleep time is a quantitative variable.

[STEP 3] : check if model explains enough variability

The p-value of overall F test is 4.043e-06, which is smaller than 0.05. Therefore, we reject the null hypothesis, and have sufficient evidence to conclude that at least one $beta_j$ does not equal to 0, and the regression is significant.

The adjusted R^2 being 0.4064 indicates that 40.64% of the variability in total sleep time is explained by the linear relationship with max life span and predation index. We think it is enough since obviously sleep time related to many other factors.

```
summary(forward)
```

```
##
## Call:
## lm(formula = total_sleep ~ max_life_span + predation_index, data = sleep2)
```

```
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1695 -2.4227 -0.3841  1.9101  7.1238
## 
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        15.4471     1.3559  11.392 7.51e-15 ***
## max_life_span      -0.1113     0.0284  -3.918 0.000301 ***
## predation_index2   -0.8986     1.5247  -0.589 0.558564
## predation_index3   -4.8357     1.6899  -2.862 0.006374 **
## predation_index4   -3.0913     1.7454  -1.771 0.083306 .
## predation_index5   -6.6951     1.5361  -4.358 7.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 3.605 on 45 degrees of freedom
## Multiple R-squared:  0.4658, Adjusted R-squared:  0.4064
## F-statistic: 7.848 on 5 and 45 DF,  p-value: 2.222e-05
```

[STEP 4] : check how each predictor impacts sleep time in the presence of other variables

The t-tests on $beta_0$, $beta_{max_life_span}$, and $beta_{predation}$ from index 2 to 5 are 7.51e-15, 0.000301, 0.558564, 0.006374, 0.083306 and 7.51e-05 respectively. All predictors, except $beta_{predation2}$ and $beta_{predation4}$, are significant.

Inter the model

$$y_{hat} = 15.4471 - 0.1113 * x_{max_life_span} - 0.8986 * x_{predation2} - 4.8357 * x_{predation3} - 3.0913 * x_{predation4} - 6.6951 * x_{predation5}$$

:

For 1 year increase in max life span, we expect the total sleep time to decrease by 0.1113 hours, holding predation index constant.

$\longrightarrow$ When $x_{predation2} = x_{predation3} = x_{predation4} = x_{predation5} = 0$ For animals with predation index 1 (least likely to be preyed upon), who has 0 years life span, we expect their total sleep time to be 15.4471 hours.

$\longrightarrow$ when $x_{predation2} = 1$, and $x_{predation3} = x_{predation4} = x_{predation5} = 0$ For animals with predation index 2, who has 0 years life span, we expect their total sleep time to be 14.5485 hours.

$\longrightarrow$ when $x_{predation3} = 1$, and $x_{predation2} = x_{predation4} = x_{predation5} = 0$ For animals with predation index 3, who has 0 years life span, we expect their total sleep time to be 10.6214 hours.

$\longrightarrow$ when $x_{predation4} = 1$, and $x_{predation2} = x_{predation3} = x_{predation5} = 0$ For animals with predation index 4, who has 0 years life span, we expect their total sleep time to be 12.3558 hours.

$\longrightarrow$ when $x_{predation5} = 1$, and $x_{predation2} = x_{predation3} = x_{predation4} = 0$ For animals with predation index 5, who has 0 years life span, we expect their total sleep time to be 8.752 hours.

```
summary(forward)
```

```
## 
## Call:
## lm(formula = total_sleep ~ max_life_span + predation_index, data = sleep2)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -7.1695 -2.4227 -0.3841  1.9101  7.1238
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       15.4471     1.3559  11.392 7.51e-15 ***
## max_life_span     -0.1113     0.0284  -3.918 0.000301 ***
## predation_index2  -0.8986     1.5247  -0.589 0.558564
## predation_index3  -4.8357     1.6899  -2.862 0.006374 **
## predation_index4  -3.0913     1.7454  -1.771 0.083306 .
## predation_index5  -6.6951     1.5361  -4.358 7.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.605 on 45 degrees of freedom
## Multiple R-squared:  0.4658, Adjusted R-squared:  0.4064
## F-statistic: 7.848 on 5 and 45 DF,  p-value: 2.222e-05
```

Part 4: summarization